

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2001, 15:22:10 ; Search time 19.89 seconds
(without alignments)
528.810 Million cell updates/sec

Title: US-09-382-242-33

Perfect score: 942
Sequence: 1 MSLNKSHWMDMIFILSF.....AVSFSLPFYIRDFKSTKR 184

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0401:*

- 1: /SIDS6/gcgdata/geneseq/geneseq/AA1980.DAT:*
- 2: /SIDS6/gcgdata/geneseq/geneseq/AA1981.DAT:*
- 3: /SIDS6/gcgdata/geneseq/geneseq/AA1982.DAT:*
- 4: /SIDS6/gcgdata/geneseq/geneseq/AA1983.DAT:*
- 5: /SIDS6/gcgdata/geneseq/geneseq/AA1984.DAT:*
- 6: /SIDS6/gcgdata/geneseq/geneseq/AA1985.DAT:*
- 7: /SIDS6/gcgdata/geneseq/geneseq/AA1986.DAT:*
- 8: /SIDS6/gcgdata/geneseq/geneseq/AA1987.DAT:*
- 9: /SIDS6/gcgdata/geneseq/geneseq/AA1988.DAT:*
- 10: /SIDS6/gcgdata/geneseq/geneseq/AA1989.DAT:*
- 11: /SIDS6/gcgdata/geneseq/geneseq/AA1990.DAT:*
- 12: /SIDS6/gcgdata/geneseq/geneseq/AA1991.DAT:*
- 13: /SIDS6/gcgdata/geneseq/geneseq/AA1992.DAT:*
- 14: /SIDS6/gcgdata/geneseq/geneseq/AA1993.DAT:*
- 15: /SIDS6/gcgdata/geneseq/geneseq/AA1994.DAT:*
- 16: /SIDS6/gcgdata/geneseq/geneseq/AA1995.DAT:*
- 17: /SIDS6/gcgdata/geneseq/geneseq/AA1996.DAT:*
- 18: /SIDS6/gcgdata/geneseq/geneseq/AA1997.DAT:*
- 19: /SIDS6/gcgdata/geneseq/geneseq/AA1998.DAT:*
- 20: /SIDS6/gcgdata/geneseq/geneseq/AA1999.DAT:*
- 21: /SIDS6/gcgdata/geneseq/geneseq/AA2000.DAT:*
- 22: /SIDS6/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	942	100.0	184	18	W23069
2	90.5	9.6	442	19	W98870
3	90.5	9.6	442	19	W1557
4	89.5	9.5	258	21	G14660
5	89.5	9.5	258	21	G14825
6	89.5	9.5	342	21	G14659
7	89.5	9.5	342	21	G14824
8	89.5	9.5	356	21	G14823
9	89.5	9.5	485	20	Y35687
10	87	9.2	249	19	W79385
11	87	9.2	412	21	Y70156

12	87	9.2	435	19	W98458	H. pylori GHPO 710
13	86.5	9.2	214	21	G14661	Arabidopsis thalia
14	86.5	9.2	397	21	G34689	Arabidopsis thalia
15	86	9.1	428	20	Y35302	Amino acid sequenc
16	86	9.1	432	18	W14081	S.thermophilus exo
17	86	9.1	473	18	W22180	S.thermophilus exo
18	85.5	9.1	212	21	G50404	Arabidopsis thalia
19	85.5	9.1	240	21	G50403	Arabidopsis thalia
20	85.5	9.1	266	21	G50402	Arabidopsis thalia
21	85.5	9.1	358	20	Y36803	Arabidopsis thalia
22	85	9.0	279	21	Y88627	Chlamydia trachoma
23	84.5	9.0	212	21	G11179	H. influenzae dime
24	84.5	9.0	240	21	G11178	Arabidopsis thalia
25	84.5	9.0	266	21	G11177	Arabidopsis thalia
26	84.5	9.0	284	21	G36490	Arabidopsis thalia
27	84.5	9.0	397	16	W13742	urea transporter p
28	84.5	9.0	430	20	W88344	Salmonella enteric
29	84	8.9	363	19	W57588	Staphylococcus aur
30	83.5	8.9	406	21	Y81562	Streptococcus pneu
31	83.5	8.9	649	21	B12140	Hydrophobic domain
32	83	8.8	335	20	W92951	MO9905287 Seg ID 1
33	83	8.8	360	16	R80953	Recombinant high a
34	82.5	8.8	378	21	B53159	Maccsa mutatta rna
35	82.5	8.8	474	19	W58862	T. halophilus xyl
36	82	8.7	840	21	B11037	S. aureus mprf pro
37	81.5	8.7	355	21	Y75278	Neisseria meningit
38	81.5	8.7	487	12	R14149	Human alpha 2 beta
39	81.5	8.7	487	18	W11804	Human alpha-2b adr
40	81.5	8.7	1137	22	B36936	Integrin alpha cha
41	81.5	8.7	1141	21	Y41752	Human PRO768 prote
42	81.5	8.7	1141	21	B44308	Human PRO768 (UNQ4
43	81	8.6	395	21	Y81541	Streptococcus pneu
44	80	8.5	504	21	Y81574	Streptococcus pneu
45	79	8.4	421	18	W20180	H. pylori putative

ALIGNMENTS

RESULT 1	
W23069	W23069 standard; Protein; 184 AA.
AC	W23069;
XX	17-FEB-1998 (first entry)
DE	Staphylothermus marinus esterase Fl-12LC.
XX	
KW	Esterase; thermostable enzyme; ester; chiral compound; cheese;
KW	pulp; paper; lignin removal; sugar; lignocellulose;
KW	disease resistance; feedstuff.
OS	Staphylothermus marinus strain Fl.
XX	
FT	Key
FT	Misc-difference 176
FT	Location/Qualifiers
XX	/note="encoded by CAG"
PN	W09730160-AL.
PD	21-AUG-1997.
XX	
PF	11-FEB-1997; 97WO-US02039.
XX	
PR	16-FEB-1996; 96US-0602359.
XX	
PA	(RECO-) RECOMBINANT BIOCATALYSIS INC.
XX	
PI	Callen W, Kosmotka A, Link S, Maffia AM, Murphy D;
PI	Reid J, Robertson DE, Swanson RV, Warren PV;
XX	
DR	WPI; 1997-425035/39.

PR 29-JUL-1997; 97US-0902615.

PR 01-APR-1997; 97US-0831309.

PR 01-APR-1997; 97US-0833457.
PR 01-APR-1997; 97US-0834705.
PR 24-JUN-1997; 97US-0881227.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI Al-Garawi A, Haas R, Kleantous H, Meyer T, Miller C;
PI Odenbreit S, Tomb J;
DR WPI; 1998-297855/26.
XX N-PSDB; V52092.
XX
PT Helicobacter polynucleotide and polypeptide sequences - useful to
PT treat or prevent gastrointestinal infection
PS Claim 1; Page 339-340; 362pp; English.
XX
XX This claimed Helicobacter pylori polypeptide, designated GHP 1719,
CC can be used in vaccination methods for preventing or treating
CC Helicobacter infection. 85 Helicobacter polypeptides (see
CC W714/4-W715/8) are claimed, as well as isolated polynucleotides
CC (see V52009-93) that encode them. The invention also provides:
CC methods for producing these Helicobacter polypeptides in
CC recombinant host systems, and related expression cassettes, vectors
CC and transformed or transfected host cells; live vaccine vectors
CC that contain the polynucleotides of the invention and which can be
CC used to prevent or treat Helicobacter infection; therapeutic and/or
CC prophylactic methods involving administration of polynucleotide
CC molecules, polypeptides or monospecific antibodies; methods for
CC detecting the presence of Helicobacter in samples using e.g.
CC the polypeptides or monospecific antibodies; and methods for
CC purifying the polypeptides by antibody-based affinity
CC chromatography.
XX
SQ Sequence 442 AA;

Query Match 9.68; Score 90.5; DB 19; Length 442;
Best Local Similarity 22.18; Pred. No. 0.027;
Matches 45; Conservative 46; Mismatches 54; Indels 59; Gaps 11;

QY 18 FSPPLTMAIAISMSSWFINMNNALSD-LCHAVKSSVAPFNLGAIIGLIVVGRNL 76
DB 211 ftslqgvtfsisglsintlyaaavtkqnlksti-----wvtsqllstsvaglmj 264
QY 77 YSMRSVKGSLLISMG--VFLNLGVPEDEVYGMIFLVSVLFSLITAYFI-AISLDR 131
DB 265 ftfvfyega-nvsqgtqlfslpvtfgqm-gaigllvsllflalafagltstvallep 322
QY 132 S-----W--IAVLLIIGHIAMWYH-----FASE- 153
DB 323 svmyltertygysrfkvwtgwjvallfvvgvllisihkdykdytlfeksldfwldfasst 382
QY 154 --IPRGAIPPELLAVFSLPFYIR 175
DB 383 lmplyg-----matflmgwvfk 401

RESULT 4
G14660
ID G14660 standard; Protein; 258 AA.
XX
AC G14660;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 14607.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0135124.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.

DE Arabidopsis thaliana protein fragment SEQ ID NO: 14834.
XX
XX Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
PN EPI03405-A2.
XX
XX 06-SEP-2000.
PD
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140595.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.

PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-015659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159285.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 9.5%; Score 89.5; DB 21; Length 258;

Best Local Similarity 23.3%; Pred. No. 0.018; Mismatches 69; Indels 51; Gaps 11;

Matches 48; Conservative 38; Mismatches 69; Indels 51; Gaps 11;

QY 1 MSLLKSHMMDIITLSEFPLTMAIA-----ISMSWF--NIMNNAISDL 45
DB 43 lamfhfwlsfflinly-yk-lp-lkqdr-layeyv-g-lwhly-g-lsmmswfw-savfnsrdvdl 101
QY 46 GHAVKSSVAPFENGLIGLIIYVGLRNL---YSMSRVGSLIISGVFLNL--IGVF 100
DB 102 terldysasa-----valglstlla-lrfdlt-veaarvmwsapila-fvthlylnfy 155
QY 101 DEYVGWTHFLVSYLEFSL---IIAFYTAISILDKSW--IAVLLIIGHTAMWYIHFASEIPR 156
DB 156 klidygw-nmlvcvmgmsq-lflwarwaavshpsnwk-lwvviagglam----- 203
QY 157 GAALPELLAVSFLEPFYIRDFYKSYT 182
DB 204 -----lleiydfppy--egyfaahs 221

RESULT 6
GI4659
ID GI4659 standard; Protein; 342 AA.

XX AC GI4659;
XX 17-OCT-2000 (first entry)
DE DE Arabidopsis thaliana protein fragment SEQ ID NO: 14606.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX EPI033405-A2.
PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123588.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135659.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.

PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151338.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156599.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160860.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match

9.5%; Score 89.5; DB 21; Length 342;

Best Local Similarity 23.3%; Pred. No. 0.025; Matches 48; Conservative 38; Mismatches 69; Indels 51; Gaps 11;

QY 1 MSLNKSHMDMIIFILSEFPLTMIALA-----ISSMSWF--NIMNNALSDL 45
Db 127 lamfhbgwlsfflamly-ykplpkqdratayeyvgjwhlygllamsvfwsvfhsrtdvl 185
QY 46 CHAVKSSVAPIFNLGILIGLIVIVGLRNL---YSMSRKSGSLISMGVFLNL--IGVF 100
Db 186 terldysa-----valldfsllalrltfdlveaarvmvsapilaftythlylnfy 239
QY 101 DEYVGWHLFLVSVLFSL---IIAYFTAISLDRKS--IAVLLIGHIAMWYLFHASEIPR 156
Db 240 klidyw--nmilvcvamyvsqflfwarwaavshpsnmkllwvvlagglam----- 287
QY 157 GAAPPELLAVSFPLPYRDFYKSYT 182
Db 288 -----llelydfpy--egyfdahs 305

RESULT 7
ID G14824
AC G14824;
DT 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 14833.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
PD
XX 25-FEB-2000; 2000EP-0301439.
PF
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137724.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140921.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.

PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	14-OCT-1999;	99US-0159639.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

```

Query Match 9.5% Score 89.5 Db 21 Length 342:
Best Local Similarity 23.3% Pred No. 0.025:
Matches 48; Conservative 38; Mismatches 69; Indels 51; Gaps 11

OY 1 MSLNRKSMWMDIIFLTSFPLTMALA-----ISM5WF--N1WNNALSDL 45
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 127 lamthgyslffimay-ykplrkgdrtaeyevglwhlyglismswsfavsfhstdvl 185
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

OY 46 GHAVSSVAPRPNCLAGLIGLIVGRLNL---YSMSVSKSLIISMGVFNLN--IGVF 100
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 186 terlgyssa-----valigfsliaaltffirveaarvmwsapilaftvthilyinfy 239
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

OY 101 DEVYGWIFLVSVLEFLS---IIAVFIALSIIDKSM-IALLVIIIGHIAMWYLIHFASEI 156
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

Db	240	kIdgW-nmIcVamgvsqILlwaarwaavshpsnWklwvviagJlam-----	287
Oy	157	GAIPPELLAVSFLEPYIRDFKSYT 182	
		:: ::	
Db	288	-----lleYdfpy--egYfdaHs 305	
RESULT	8		
GI4823			
XX	GI4823	standard; Protein; 356 AA.	
XX	AC	GI4823;	
XX	DT	17-OCT-2000 (first entry)	
XX	DE	Arabidopsis thaliana protein fragment SEQ ID NO: 14832.	
XX	KM	Protein identification; signal transduction pathway; metabolic pathway;	
XX	KM	hybridisation assay; genetic mapping; gene expression control; promoter;	
XX	OS	termination sequence.	
XX	OS	Arabidopsis thaliana.	
PN		EP1033405-A2.	
XX	PD	06-SEP-2000.	
PF		25-FEB-2000; 2000EP-0301439.	
XX			
PR	25-FEB-1999;	99US-0121825.	
PR	05-MAR-1999;	99US-0123180.	
PR	09-MAR-1999;	99US-0123548.	
PR	23-MAR-1999;	99US-0125788.	
PR	25-MAR-1999;	99US-0126264.	
PR	29-MAR-1999;	99US-0126785.	
PR	01-APR-1999;	99US-0127462.	
PR	06-APR-1999;	99US-0128234.	
PR	08-APR-1999;	99US-0128714.	
PR	16-APR-1999;	99US-0129845.	
PR	19-APR-1999;	99US-0130077.	
PR	21-APR-1999;	99US-0130449.	
PR	23-APR-1999;	99US-0130510.	
PR	28-APR-1999;	99US-0130891.	
PR	28-APR-1999;	99US-0131449.	
PR	30-APR-1999;	99US-0132048.	
PR	30-APR-1999;	99US-0132407.	
PR	04-MAY-1999;	99US-0132484.	
PR	05-MAY-1999;	99US-0132485.	
PR	06-MAY-1999;	99US-0132486.	
PR	06-MAY-1999;	99US-0132487.	
PR	07-MAY-1999;	99US-0132487.	
PR	11-MAY-1999;	99US-0134256.	
PR	14-MAY-1999;	99US-0134218.	
PR	14-MAY-1999;	99US-0134219.	
PR	14-MAY-1999;	99US-0134370.	
PR	14-MAY-1999;	99US-0134370.	
PR	18-MAY-1999;	99US-0134768.	
PR	19-MAY-1999;	99US-0134941.	
PR	20-MAY-1999;	99US-0135124.	
PR	21-MAY-1999;	99US-0135353.	
PR	24-MAY-1999;	99US-0135629.	
PR	25-MAY-1999;	99US-0136021.	
PR	27-MAY-1999;	99US-0136392.	
PR	28-MAY-1999;	99US-0136782.	
PR	01-JUN-1999;	99US-0137222.	
PR	03-JUN-1999;	99US-0137528.	
PR	04-JUN-1999;	99US-0137502.	
PR	07-JUN-1999;	99US-0137724.	
PR	08-JUN-1999;	99US-0138094.	
PR	10-JUN-1999;	99US-0138540.	
PR	10-JUN-1999;	99US-0138847.	
PR	14-JUN-1999;	99US-0139119.	
PR	16-JUN-1999;	99US-0139452.	

PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144864.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145097.
PR 22-JUL-1999; 99US-0145112.
PR 22-JUL-1999; 99US-0145115.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149920.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159684.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 23-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 9.58; Score 89.5; DB 21; Length 356;
Best Local Similarity 23.3%; Pred. No. 0.027;
Matches 48; Conservative 38; Mismatches 69; Indels 51; Gaps 11;

QY 1 MSLNKSHMDMIIIFILSPSPITMIALA-----ISMSMF--NIMNNAISDL 45
Db 141 IamfhgwistfImly-yk!p!kqdr!cayeygw!whlyg!lsmns!wfa!v!ns!d!v!l 199
QY 46 GHAVKSSVAP!FNGLA!IGL!IV!GLRNL---YSMSRVKGS!L!ISM!GVFLNL--IGVF 100


```

Oy 52 --SVAPIFNLGAI--GGIIE--VYGLRNLYSMSRVKGLSIISMGVFLNLIGVPEVVGW 106
    |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 101 lKstgiacIcdnllgglltsaItcWlnhrryskrlp-----emvgvIdg---g 145
Oy 107 IHFLVSVLEFSLIIVFAISILDKSMIAV-----LLIIGHIAMYLHFASEIIPR 156
    |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 146 lftvrlstfslfmpIaaItcvl-----wptvghglsqmghflIasgyIgwlyhflerv-- 199
Oy 157 GAAPPELLAVFSELPFY 174
Db 200 --lIptglnhIfIyapIev 215

RESULT 11
Y70156
ID Y70156 standard; Protein: 412 AA.
XX Y70156;
XX
XX Y70156;
XX
XX 06-JUN-2000 (first entry)
XX
XX Staphylococcus aureus protein (168339_1).
XX
XX Vaccine: antibacterial; prevention; attenuation; detection;
XX Staphylococcal infection; neonatal conjunctivitis; skin infection;
XX toxic shock syndrome; osteomyelitis.
XX
XX Staphylococcus aureus.
XX
XX WO200012678-A2.
XX
XX 09-MAR-2000.
XX
XX 31-AUG-1999; 99WO-US19726.
XX
XX 01-SEP-1998; 98US-0098964.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Bailey CC, Choi GH;
XX
XX WPI: 2000-237864/20.
XX DR N-PSDB; Z51236, Z51240.
XX
XX Staphylococcus aureus polypeptide useful for preventing or attenuating
XX a Staphylococcal infection comprises one of 32 sequences of 100-1277
XX amino acids or their fragments -
XX
XX Claim 9; Page 22; 144p; English.
XX
XX The present sequence is a protein (168339_1)
XX from Staphylococcus aureus genomic DNA library.
XX This sequence is useful in preparation of vaccines for
XX prevention or attenuation of Staphylococcal infections (especially
XX S.aureus infections) which may cause conditions such as neonatal
XX conjunctivitis, osteomyelitis, skin infections and toxic shock syndrome.
XX The present sequence is also useful for detecting Staphylococcal
XX infections in biological samples.
XX
XX Sequence 412 AA:

Query Match 9.2%; Score 87; DB 21; Length 412;
Best Local Similarity 20.7%; Pred. No. 0.063; Mismatches 30; Indels 36; Gaps 5;
Matches 25; Conservative 30;

Oy 37 lWNNALSDLGAAVKSVAPlFNlGAlGlllVlVlGRLNLYSMRSRVKGLSIISMGVFLNL 96
    |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 290 wlnals-----vIk-----YclfgvgvIgdvYvhlg-----sqInglllvahntYql 332
    |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Oy 97 lGVPEVVG--WIHFLVSVL-----FLSIIVFAISILDKSMIAVL 137
    |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|

```

```

Db      333 faewngllfgalfflfnllyllfelffrnlsqknvtaivmltmlyltvsnfmsryvaf 392
QY      138 L 138
Db      393 I 393

RESULT 12
W98458
ID      W98458 standard; Protein; 435 AA.
XX
AC      W98458;
XX
DT      31-MAR-1999 (first entry)
XX
DE      H. pylori GHPO 710 protein.
XX
KM      GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW      peptic ulcer disease.
XX
OS      Helicobacter pylori.
XX
PN      WO9843478-A1.
XX
PD      08-OCT-1998.
XX
PF      01-APR-1998; 98WO-US06371.
XX
PR      29-JUL-1997; 97US-0902615.
PR      01-APR-1997; 97US-0833457.
PR      24-JUN-1997; 97US-0881227.
XX
PA      (HUMA-) HUMAN GENOME SCI INC.
XX      (INMR-) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI      Al-garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;
DR      WPI: 1998-542293/46.
DR      N-PSDB: X14177.
XX
PT      New isolated Helicobacter polynucleotides - used to develop products
PT      for the diagnosis, prevention and treatment of Helicobacter
PT      infections and gastrointestinal diseases
PS      Claim 8; Page 805-807; 2054pp; English.
XX
CC      This sequence represents a Helicobacter pylori GHPO protein of the
CC      invention. The polypeptides can be used for preventing or treating
CC      Helicobacter infections, and gastroduodenal diseases associated with
CC      these infections, including acute, chronic, and atrophic gastritis, and
CC      peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC      used for the production of antibodies. The products can also be used for
CC      detection and diagnosis.
XX
SQ      Sequence 435 AA:

Query Match          9.2%; Score 87; DB 19; Length 435;
Best Local Similarity 18.2%; Pred. No. 0.067;
Matches 40; Conservative 39; Mismatches 73; Indels 66; Gaps 7

QY      22 LTMIALAISMSMFN-INNNALSDLGHAVKSSVAPIFNIGL-----AIGG 65
DB      49 lsaatllflsvfnglwann--tplamsvglslayfsfglvqgklpwgsalgalvalsg 106
QY      66 IIVIVGLRNLYSW-----SRVKSGLIISMGVFLNLIGVFEDVYGVHIFLVS----- 112
DB      107 aifvllstfkfsvwmrslpsqlrravsglgaflafllike-----mlivvltlxatiyt 161
QY      113 -----VLFELSLINAFIALISLIDKSWIIVALLIGHIIMWYLIHFASERGAIPRL 163
DB      162 lggdfgdphvllvgvllltfalfayltklqgsfliaavlltsllawvklapyseelfsmpas 221

```

QY	164	LA VSF-----LPFYRDFKS 160
DB	222	Igpiatqldfkylffdasgaftalypviltffvtoldfs 261
RESULT	13	
ID	G14661	
AC	G14661 standard. Protein; 214 AA.	
XX	G14661;	
DT	17-OCT-2000 (first entry)	
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 14608.	
KW	Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.	
XX	Arabidopsis thaliana.	
OS	EP1033405-A2.	
PN	EP1033405-A2.	
XX	06-SEP-2000.	
PD	06-SEP-2000.	
XX	25-FEB-2000; 2000EP-0301439.	
PF	25-FEB-2000; 2000EP-0301439.	
XX	25-FEB-1999; 9905-0121825.	
PR	05-MAR-1999; 9905-0123180.	
PR	09-MAR-1999; 9905-0123548.	
PR	23-MAR-1999; 9905-0125788.	
PR	25-MAR-1999; 9905-0126264.	
PR	29-MAR-1999; 9905-0126785.	
PR	01-APR-1999; 9905-0127462.	
PR	06-APR-1999; 9905-0128234.	
PR	08-APR-1999; 9905-0128714.	
PR	16-APR-1999; 9905-0129845.	
PR	19-APR-1999; 9905-0130077.	
PR	21-APR-1999; 9905-0130449.	
PR	23-APR-1999; 9905-0130510.	
PR	28-APR-1999; 9905-0130891.	
PR	30-APR-1999; 9905-0131449.	
PR	30-APR-1999; 9905-0132048.	
PR	04-MAY-1999; 9905-0132407.	
PR	05-MAY-1999; 9905-0132484.	
PR	06-MAY-1999; 9905-0132485.	
PR	06-MAY-1999; 9905-0132486.	
PR	07-MAY-1999; 9905-0132487.	
PR	11-MAY-1999; 9905-0132863.	
PR	14-MAY-1999; 9905-0134256.	
PR	14-MAY-1999; 9905-0134218.	
PR	14-MAY-1999; 9905-0134219.	
PR	14-MAY-1999; 9905-0134370.	
PR	18-MAY-1999; 9905-0134768.	
PR	19-MAY-1999; 9905-0134941.	
PR	20-MAY-1999; 9905-0135124.	
PR	21-MAY-1999; 9905-0135353.	
PR	24-MAY-1999; 9905-0135629.	
PR	25-MAY-1999; 9905-0136021.	
PR	27-MAY-1999; 9905-0136392.	
PR	28-MAY-1999; 9905-0136782.	
PR	01-JUN-1999; 9905-0137222.	
PR	03-JUN-1999; 9905-0137528.	
PR	04-JUN-1999; 9905-0137502.	
PR	07-JUN-1999; 9905-0137724.	
PR	08-JUN-1999; 9905-0138094.	
PR	10-JUN-1999; 9905-0138540.	
PR	10-JUN-1999; 9905-0138847.	
PR	14-JUN-1999; 9905-0139119.	
PR	16-JUN-1999; 9905-0139452.	
PR	16-JUN-1999; 9905-0139453.	
PR	17-JUN-1999; 9905-0139492.	

```
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 25-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159684.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160815.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160768.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 26-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161992.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

Query Match 9.2%; Score 86.5; DB 21; Length 214;
Best Local Similarity 23.9%; Pred. No. 0.031;
Matches 48; Conservative 34; Mismatches 68; Indels 51; Gaps 11;
```

```
QY 6 HSNWDMITFILSFSPITMIALA-----ISMSSWF--NIMNNAISDLGHAVK 50
DB 4 hgvistfflmly-ykplpkqdtayeyvglwhlyglismswfswsavfhsrddvltterld 62
QY 51 SSVAPIRNLGLAIGLIVVGLNL---YSMSRVKSLIISKVFNL--IGVFDEYVG 105
DB 63 yssa-----vallgfsllaliltfdrveaarvmwsaplatfvtchlyinfykidyg 116
```

```
QY 106 WHPFLVLEFLS---IIAFLAISIDKSM-IAVLLIIGHIMWYLFHFASEIPRGAAP 161
DB 117 W-mivcvmavgsqglfwatwaavsshpnmkllwvvlagglam----- 159
QY 162 ELAAVFSFLPPYIRDRYKSYT 182
DB 160 -lleiyafppy--egyfdans 177

RESULT 14
G36489
ID G36489 standard; Protein; 397 AA.
XX
AC G36489;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 44725.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 9905-0121825.
PR 05-MAR-1999; 9905-0123180.
PR 09-MAR-1999; 9905-0123548.
PR 23-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126264.
PR 29-MAR-1999; 9905-0126785.
PR 01-APR-1999; 9905-0127462.
PR 06-APR-1999; 9905-0128234.
PR 08-APR-1999; 9905-0128714.
PR 16-APR-1999; 9905-0129845.
PR 19-APR-1999; 9905-0130077.
PR 21-APR-1999; 9905-0130449.
PR 23-APR-1999; 9905-0130891.
PR 28-APR-1999; 9905-0131449.
PR 30-APR-1999; 9905-0132048.
PR 30-APR-1999; 9905-0132407.
PR 04-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 06-MAY-1999; 9905-0132487.
PR 07-MAY-1999; 9905-0132863.
PR 11-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134221.
PR 14-MAY-1999; 9905-0134370.
PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136382.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
```


Db 105 vlejllfamwylnfyfnkykalkahapmvtlvqf---avgsvllttmwnly 160
 QY 78 SMSRVKG---SLIISMGVFLNLIGVEDEV-----YGMHFLVSLFPLSIIVARIASI 128
 Db 161 krpklsgaqllaillplavhclgnlftmsjgkvsstfhtkameplfsvllsamflge 220
 QY 129 LDKSWI--AVLLIGHIAM 145
 Db 221 ktpwvlgelivpivgval 239

RESULT 15

Y35302
 ID Y35302 standard; Protein; 428 AA.

XX Y35302;

DT 13-SEP-1999 (first entry)

DE Amino acid sequence of a Chlamydia pneumoniae protein.

XX Respiratory disease: pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope.

XX Chlamydia pneumoniae.

XX OS MO9927105-A2.

XX PN 03-JUN-1999.

XX PF 20-NOV-1998; 98WO-IB01890.

XX PR 04-NOV-1998; 98US-0107078.

XX PR 21-NOV-1997; 97FR-0014673.

XX (GEST) GENSET.

XX Griffais R;

XX WPI; 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae

XX Page 1119; Disclosure: 1912pp; English.

CC Y34584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see X91990) of Chlamydia pneumoniae.

CC C. pneumoniae causes respiratory disease such as pneumonia and

CC bronchitis and is thought to be a contributing factor in heart

CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema

CC nodosum or pharyngitis. The polypeptides encoded by the open reading

CC frames of the C. pneumoniae genome (see Y34584-Y35879) can be used in

CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae

CC nucleotide sequences can also be used as immunogenic compositions,

CC especially where the vector directs the expression of a neutralising

CC epitope of C. pneumoniae.

CC SQ Sequence 428 AA;

Query Match 9.1%; Score 86; DB 20; Length 428;

Best Local Similarity 23.1%; Pred. No. 0.086;

Matches 48; Conservative 30; Mismatches 56; Indels 74; Gaps 11;

QY 6 HSMWDMIFILISFPLTMLALISMSSWFNIMNALSGLGHAVKSV-----53

Db 1 ysmpllllflv-----lsgfycwnlgandvanavgsvsgvllrlgavv 46

QY 54 -APIFN-----LGLAIGIL---IVIVGLRNLYSMSRVKG--SLIISMGVFLNLIGVED 101

Db 47 laaifefgalllgdrvagtllessivsvtnpmlasgdyymgtaalllatgywlglaeff- 105

QY 102 EYVGW----IHFLVLPFLSI-----IAYFIAISLDKSMIANVLLIGHIAMWYLHFA 151
 Db 106 ---gwpvsthsivgavlgfvlvgkgtliynsvglllswllspfmngcva---yl--- 157
 QY 152 SEIPRGAALPELLAVFSFLP---FYIRD 176
 Db 158 -----lfsflrrhlffyknd 171

Search completed: May 30, 2001, 15:23:49
 Job time: 99 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2001, 15:22:30 ; Search time 11.2 Seconds
(Without alignments)
315.607 Million cell updates/sec

Title: US-09-382-242-33

Perfect score: 942
Sequence: 1 MSLNKSHWMDMIFILSFSE.....AVFSFLPFYIRDFKSYTKR 184

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PT05_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	942	100.0	184	2	US-08-602-359A-33
2	86	9.1	473	1	US-08-597-236-13
3	86	9.1	473	1	US-08-746-682A-13
4	85	9.0	279	3	US-08-549-515-7
5	84.5	9.0	397	1	US-08-098-141-2
6	84	8.9	363	4	US-09-046-086-2
7	78.5	8.3	330	1	US-08-118-270-20
8	78.5	8.3	330	5	PCT-US93-08528-20
9	77	8.2	266	2	US-08-225-480-7
10	77	8.2	266	2	US-09-097-759-4
11	77	8.2	299	2	US-09-097-759-3
12	77	8.2	306	2	US-09-097-759-2
13	77	8.2	341	1	US-08-423-564-5
14	76.5	8.1	394	2	US-08-902-853-1
15	76.5	8.1	461	1	US-08-194-338-4
16	76	8.1	330	1	US-08-118-270-19
17	76	8.1	330	5	PCT-US93-08528-19
18	76	8.1	355	1	US-07-759-568-1
19	76	8.1	355	1	US-08-450-393A-8
20	76	8.1	355	2	US-08-390-000A-5
21	76	8.1	355	5	PCT-US95-00476-8
22	76	8.1	355	5	PCT-US95-00476-8
23	76	8.1	360	1	US-08-202-056-7
24	75	8.0	1531	1	US-08-141-893-2
25	75	8.0	1531	1	US-08-463-092B-2
26	75	8.0	1531	1	US-08-463-092B-4
27	75	8.0	1531	2	US-08-462-109A-2

28	75	8.0	1531	2	US-08-462-109A-4	Sequence 4, Appl1
29	75	8.0	1531	2	US-08-460-907B-2	Sequence 2, Appl1
30	75	8.0	1531	2	US-08-460-907B-4	Sequence 4, Appl1
31	75	8.0	1531	3	US-08-463-179A-2	Sequence 2, Appl1
32	75	8.0	1531	3	US-08-463-179A-2	Sequence 4, Appl1
33	75	8.0	1531	3	US-08-461-384B-2	Sequence 2, Appl1
34	75	8.0	1531	3	US-08-461-384B-4	Sequence 4, Appl1
35	75	8.0	1531	3	US-08-407-207A-2	Sequence 2, Appl1
36	74	7.9	458	1	US-08-310-271-2	Sequence 2, Appl1
37	74	7.9	458	3	US-09-292-071-27	Sequence 21, Appl1
38	74	7.9	458	3	US-09-292-071-29	Sequence 29, Appl1
39	74	7.9	458	4	US-09-292-069A-27	Sequence 21, Appl1
40	74	7.9	458	4	US-09-292-069A-29	Sequence 29, Appl1
41	74	7.9	482	4	US-08-637-823B-32	Sequence 32, Appl1
42	73.5	7.8	652	1	US-08-050-684-2	Sequence 2, Appl1
43	73.5	7.8	652	1	US-08-582-719-2	Sequence 2, Appl1
44	73	7.7	341	2	US-08-846-762-92	Sequence 92, Appl1
45	73	7.7	1090	3	US-08-307-896-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-08-602-359A-33
Sequence 33, Application US/08602359A
Patent No. 5942430
GENERAL INFORMATION:
APPLICANT: ROBERTSON, Daniel E.
APPLICANT: MURPHY, Dennis
APPLICANT: REID, John
APPLICANT: MAFRIA, Anthony
APPLICANT: LINK, Steven
APPLICANT: SMANSON, Ronald V.
APPLICANT: WARREN, Patrick V.
APPLICANT: KOSMOTKA, Anna
TITLE OF INVENTION: ESTERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & RICHARDSON P.C.
STREET: 4225 EXECUTIVE SQUARE, STE 1400
CITY: LA JOLLA
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/602,359A
FILING DATE: February 16, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: HAILE, LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070
TELEFAX: 619-678-5099
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 184 AMINO ACIDS
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-602-359A-33

Query Match 100.0%; Score 942; DB 2; Length 184;
Best Local Similarity 100.0%; Pred. No. 3.5e-89;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLNHSMWMDIIFLISFSPPLTMTALAISMSSWNNNNALSDGHAIVSVAFIFNIG 60
DB 1 MSLNHSMWMDIIFLISFSPPLTMTALAISMSSWNNNNALSDGHAIVSVAFIFNIG 60
QY 61 LAIGGILIVIGLRNLYMSRKYSGSLISMGVFNLIIGVDEYVGMHPLVSVFELSTI 120
DB 61 LAIGGILIVIGLRNLYMSRKYSGSLISMGVFNLIIGVDEYVGMHPLVSVFELSTI 120
QY 121 AVFAISILDSKSWIAVLLIIGHIAMVYLHFAISEIPRGAIPPELLAVFSFLPYIRDYKFS 180
DB 121 AVFAISILDSKSWIAVLLIIGHIAMVYLHFAISEIPRGAIPPELLAVFSFLPYIRDYKFS 180
QY 181 YTKR 184
DB 181 YTKR 184

RESULT 2
US-08-597-236-13
Sequence 13, Application US/08597236
Patent No. 5733765
GENERAL INFORMATION:
APPLICANT: STINGELE, Francesca
APPLICANT: MOLETT, Beat
TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
TITLE OF INVENTION: EXOPOLYSACCHARIDES
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americans
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/597,236
FILING DATE:
CLASSIFICATION: 426
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95201669.9
FILING DATE: 20-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fanucci A., Allan
REGISTRATION NUMBER: 30256
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 473 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-597-236-13

Query Match 9.1%; Score 86; DB 1; Length 473;
Best Local Similarity 24.5%; Pred. No. 0.36;
Matches 52; Conservative 36; Mismatches 74; Indels 50; Gaps 11;

QY 2 SLNKHSMWMDIIFLISFSPPLTMTALAISMSSWNNNNALSDGHAIV 49
DB 74 SMTKESVLTNSLVWCFSAVFCLGACITVALNLSNKWYLLALLTFNLFQGGOSTLSQYA 133
QY 50 KS-SVAPFENGLAIGLIIYVIGLRNLYMSRKYSGSLISMGVFNLIIGVDEYVGMH 108
DB 134 RGICKSKIF---AAGVILTF-----LTGALNILEFLVYLP-----GITGYLM 173

QY 109 FLV-----SVLEFSLIAVF--IAISILDSKSWI-----AVLLIIGHIAMVYLHFASE-- 153
DB 174 SLVLAVNGTILFFACTLSIMKEISFKIIDKLIWQMLYALPLIPSSILMWLNASSRYF 233
QY 154 --IPRGAIPPELLAVFSFLPYIRDYKFSYTK 183
DB 234 VLFFLGAGANGLAVATKIPSIISIFNTIFIQ 265

RESULT 3
US-08-746-682A-13
Sequence 13, Application US/08746682A
Patent No. 5786184
GENERAL INFORMATION:
APPLICANT: STINGELE, Francesca
APPLICANT: MOLETT, Beat
TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
TITLE OF INVENTION: EXOPOLYSACCHARIDES
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americans
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/746,682A
FILING DATE: 14-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/597,236
FILING DATE: 20-JUN-1995
APPLICATION NUMBER: EP 95201669.9
FILING DATE: 20-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fanucci A., Allan
REGISTRATION NUMBER: 30256
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 473 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-746-682A-13

Query Match 9.1%; Score 86; DB 1; Length 473;
Best Local Similarity 24.5%; Pred. No. 0.36;
Matches 52; Conservative 36; Mismatches 74; Indels 50; Gaps 11;

QY 2 SLNKHSMWMDIIFLISFSPPLTMTALAISMSSWNNNNALSDGHAIV 49
DB 74 SMTKESVLTNSLVWCFSAVFCLGACITVALNLSNKWYLLALLTFNLFQGGOSTLSQYA 133
QY 50 KS-SVAPFENGLAIGLIIYVIGLRNLYMSRKYSGSLISMGVFNLIIGVDEYVGMH 108
DB 134 RGICKSKIF---AAGVILTF-----LTGALNILEFLVYLP-----GITGYLM 173
QY 109 FLV-----SVLEFSLIAVF--IAISILDSKSWI-----AVLLIIGHIAMVYLHFASE-- 153
DB 174 SLVLAVNGTILFFACTLSIMKEISFKIIDKLIWQMLYALPLIPSSILMWLNASSRYF 233
QY 154 --IPRGAIPPELLAVFSFLPYIRDYKFSYTK 183
DB 234 VLFFLGAGANGLAVATKIPSIISIFNTIFIQ 265

APPLICATION NUMBER: US/09/046,086
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/043,489
FILING DATE: 10-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Falk, Stephen T
REGISTRATION NUMBER: 36,795
REFERENCE/DOCKET NUMBER: GM50025
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2488
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 363 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-046-086-2

Query Match 8.9%; Score 84; DB 4; Length 363;
Best Local Similarity 27.3%; Pred. No. 0.42;
Matches 30; Conservative 21; Mismatches 45; Indels 14; Gaps 4;

QY 82 VKSLIISMGVFLNLI---GV-FDEVYG--WIHPLVSLFELSLIAVFAISILDKSMI 134
DB 4 LKDSIAEISSILFLIFFIAGIFENEYGPWLYIISVLSLILVYNNRLNTLMF 63
QY 135 AVLLIIGHIAMIYLFHFASEIPRGAIPPELLAVSFLEPIYINDYFSYTKR 184
DB 64 YLLIHIFICYEVFSVH-----PMLSLFFYSARAVFTFRKNVKK 106

RESULT 7
US-08-118-270-20
Sequence 20, Application US/08118270
Patent No. 5508384
GENERAL INFORMATION:
APPLICANT: Murphy, Randall B.
TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
NUMBER OF SEQUENCES: 348
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/118,270
FILING DATE: 09-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/943,236
FILING DATE: 10-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, Kevin G.
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: MURPHY-2A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:
LENGTH: 330 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-118-270-20

Query Match 8.3%; Score 78.5; DB 1; Length 330;
Best Local Similarity 28.9%; Pred. No. 1.4;
Matches 33; Conservative 17; Mismatches 43; Indels 21; Gaps 5;

QY 25 IALAISMSWFNINNNALSDIGHAVKSSV-AP--IFNLGLAIGLILVYGLRNLYSMSR 81
DB 4 IAAATFLIFLFTIGNALVIITAVLSRSLRPQNLFLVSIADAILV----- 50
QY 82 VKSLIISMGVFLNLI-G--VFDEVYGIHFLVSVLFELSLIAVFAISILDKSM 133
DB 51 -ATLIIFSLANELLGWYFRRTWCEVYIALDVLECTSIHLCATS-LDRYW 101

RESULT 8
PCT-US93-08528-20
Sequence 20, Application PC/TUS9308528
GENERAL INFORMATION:
APPLICANT: New York University
TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
NUMBER OF SEQUENCES: 348
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/08528
FILING DATE: 09-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/943,236
FILING DATE: 10-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, Kevin G.
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: MURPHY-2 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 330 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US93-08528-20

Query Match 8.3%; Score 78.5; DB 5; Length 330;
Best Local Similarity 28.9%; Pred. No. 1.4;
Matches 33; Conservative 17; Mismatches 43; Indels 21; Gaps 5;

QY 25 IALAISMSWFNINNNALSDIGHAVKSSV-AP--IFNLGLAIGLILVYGLRNLYSMSR 81
DB 4 IAAATFLIFLFTIGNALVIITAVLSRSLRPQNLFLVSIADAILV----- 50

Db 189 FFVASLIDGSATMHSVLTIDMTTILSLMYLAFVART 226

RESULT 12

US-09-097-759-2

Sequence 2, Application US/09097759A

Patent No. 5972663

GENERAL INFORMATION:

APPLICANT: Winterhalter Mr., Christopher

APPLICANT: Leinfelder Mr., Manfred

TITLE OF INVENTION: Microorganisms and Processes for the Fermentative

TITLE OF INVENTION: Preparation of L-cysteine,

TITLE OF INVENTION: L-cysteine, N-acetylsuccinyl or thiazolidine Derivatives

FILE REFERENCE: Winterhalter

CURRENT APPLICATION NUMBER: US/09/097,759A

CURRENT FILING DATE: 1998-06-16

EARLIER APPLICATION NUMBER: DE 197 26 083

EARLIER FILING DATE: 1997-06-19

NUMBER OF SEQ ID NOS: 12

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 2

LENGTH: 306

TYPE: PR

ORGANISM: Escherichia coli

US-09-097-759-2

Query Match

Best Local Similarity 8.2%; Score 77; DB 2; Length 306;

Matches 38; Conservative 24; Mismatches 66; Indels 30; Gaps 6;

QY 16 LSFSEPLTMTALATISMSWENINNALSDLGHAVKSSVAPFNLGATIGLIVYGLN 75

Db 87 INFMPAGLASLVQAQAFITMGAFT-FEGERLHGK--QLAGTALMIFGLVLEDSLN 143

QY 76 LYSMSRKGLISLISMGVFLNIGVFDEYGVWTH-----FLYSVLEFLSI 121

Db 144 -----GQHVAMLGPMGLTAAAFSWACGNIFNKIKMSHSTRPVMVSLVMSALIPPI 195

QY 122 YFIAISILDKS--WIAVLLIIGH--IAMWYLHPASEI 154

Db 196 FFVASLIDGSATMHSVLTIDMTTILSLMYLAFVART 233

RESULT 13

US-08-423-564-5

Sequence 5, Application US/08423564

Patent No. 5654417

Patent No. 5654417 5654417

GENERAL INFORMATION:

APPLICANT: Tarr, Phillip I.

APPLICANT: Bilge, Sina S.

APPLICANT: Vary, Jr., James C.

TITLE OF INVENTION: NUCLEIC ACID PROBES FOR DETECTING E. coli O157:H7

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Christiansen, O'Connor, Johnson, Kindness

STREET: 1420 5th Ave., Suite 2800

CITY: Seattle

STATE: WA

COUNTRY: USA

ZIP: 98101-2347

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/423,564

FILING DATE: 14-APR-1995

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Broderick, Thomas F.

REGISTRATION NUMBER: 31,332

REFERENCE/DOCKET NUMBER: CHOR-1-8116

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 682-8100

TELEFAX: (206) 224-0779

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 341 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

DESCRIPTION: amino acids encoded by

DESCRIPTION: nucleotides 1-1026 of SEQ ID NO:1

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Escherichia coli

US-08-423-564-5

Query Match

Best Local Similarity 8.2%; Score 77; DB 1; Length 341;

Matches 43; Conservative 28; Mismatches 68; Indels 44; Gaps 9;

QY 19 SFTPLTMTALATIS-MSSW-----FNIMNNALSDLGHAVKSSVAPFNLGATIGLIVYGL 73

Db 1 SFTVTAICITTSMDIMQVLYKGINENVYAELOHSIKVPIY-IFGLTIYSNGQVYKMG 59

QY 74 RNLYMSRVKSGSLISMGVFLNIGVFDEYGVWTH-----FLYSVLEFLSI 128

Db 60 QKAVI-SNIVAVIRPILSLITLVIS-----SKRLAGLPVIVSTLGIQYISGIYLTINL 112

QY 129 LDKRWIA-----VLLIGHIAMW-LHPASEIRGAALPELLA 165

Db 113 IIRKLIFTVNHNHAKREAPYLLINGFEFFILOGLATWSGDNFIITSLGVY---YVA 169

QY 166 VFS 168

Db 170 VFS 172

RESULT 14

US-08-902-853-1

Sequence 1, Application US/08902853

Patent No. 5945330

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Corley, Neil C.

APPLICANT: Shah, Puryi

APPLICANT: Lal, Preeti

TITLE OF INVENTION: HUMAN LONGEVITY-ASSURANCE PROTEIN HOMOLOGS

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/902,853

FILING DATE: Herewith

CLASSIFICATION: ?

PRIOR APPLICATION NUMBER:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

[illegible]

```

Oy      135 AVLLIGHIAMWYLFH-----ASIEIPGAIPELLAVFSFLPEYI 174
          ||| | | | | | | | | | | | | | | | | | | | | |
Db      194 LVALLIGIIACWVLSINATNRQLTGQTFMGRIITLLCTFFPLVYFVL 242

RESULT      8
B75542
Conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: B75542
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
; M., Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zaleski, C.;
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: B75542
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-373 <WH>
A:Cross-references: GB:AE001886; GB:AE000513; NID:g6457921; PIDN:AAF09638.1; PID:g645
C:Genetics:
A:Experimental source: strain R1
C:Gene: DR0252
A:Map position: 1
C:Superfamily: Bacillus subtilis conserved hypothetical protein yuef

Query Match      11.0%; Score 104.5; DB 2; Length 373;
Best Local Similarity 27.4%; Pred. No. 0.12;
Matches 43; Conservative 29; Mismatches 44; Indels 41; Gaps 9;

Oy      30 SSMSSF-NIMNNALSDLGHAKSSVAFIFNLGLAIGT-----LIIVGLRNLYXMSRYK 83
          ::::: | | | | | | | | | | | | | | | | | | | | | |
Db      142 ALRDWIANIIGNNYPIIDNAL-SSTGTLEFNRLISIGVIGVILLILSVYLMDYARVN 200
          : | | | | | | | | | | | | | | | | | | | | | |
Oy      84 GSLLISMG-----VELNLIGVFDEYVGMIHFVSLVFLSIHAAYFAISLDKSWIA 135
          ||: | | | | | | | | | | | | | | | | | | | | |
Db      201 ASLLAAFRPMQPKLVESDLIGT--AVGVYVRQL-----LIAMFTGIFF----WLG 247
          | | | | | | | | | | | | | | | | | | | | | |
Oy      136 VLLIGHIAMWYLFHASEIPRGAIPELLAVFSFLPF 172
          | | | | | | | | | | | | | | | | | | | | | |
Db      248 -LTIVG-----IPSAAGFLGAGFNIIVPY 271
          | | | | | | | | | | | | | | | | | | | | | |

RESULT      9
AB2043
Inner membrane copper tolerance protein cycZ VC2701 [similarity] - Vibrio cholerae (g
S:Alternate names: thiol:disulfide interchange protein Dsbd
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
C:Accession: AB2043
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
; Chaudron, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
; L. R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; MUID:20406833
A:Accession: AB2043
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-600 <HEI>
A:Cross-references: GB:AE004336; GB:AE003852; NID:g9657296; PIDN:AAF95841.1; GSPDB:CN
A:Experimental source: serogroup O1; strain NI6561; biotype El Tor
C:Genetics:
A:Gene: VC2701
A:Map position: 1

Query Match      *11.0%; Score 103.5; DB 2; Length 600;
Best Local Similarity 22.5%; Pred. No. 0.22;
```

Matches 42; Conservative 40; Mismatches 82; Indels 23; Gaps 7;

OY 5 KHSWMDMIFILSFSPPLTMIAL-AISMSSWFNIMNALSOLGHAVSSVAPIFNLGLAI 63
 Db 271 OHPIVLMGLSVLEFVALISMGLSLQIPSCVQWLNLSNAQOG--GSLPGVFRMGAIIS 328
 OY 64 -----GILYIVGLRNLYMSRVKGSLLISMVGLNLIGVF-----DEVGW 106
 Db 329 GLVCSPTTAPLGSALGLVVAQSGDLTGAVLALYALAIAGMGPILLIVAFGKMLPKAGNW 388
 OY 107 IHFLVSVLFSLIITA-VFIAISILDKSWIANV-LIIGHIAM-WIHFASSELPRGAIPEL 163
 Db 389 MERKKTLEGFLLAAPILFLERIVPEEFSSVLSALGLAAGGMVHVKNSLPFGWKQSL 448
 OY 164 LAVPSFL 170
 Db 449 IGIVAIL 455

RESULT 10

LABECA

Latent membrane protein LMP1 - human herpesvirus 4 (strain CAO)

C:Species: human herpesvirus 4, Epstein-Barr virus
 C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
 C:Accession: J01434; G00065; S21660
 R:Hu, L.F.; Zbarovsky, E.R.; Chen, F.; Cao, S.L.; Ernberg, I.; Klein, G.; Winberg, G.
 J. Gen. Virol. 72, 2399-2409, 1991
 A:Title: Isolation and sequencing of the Epstein-Barr virus BNLF-1 gene (LMP1) from a CH
 A:Reference number: J01434; MUID:92013956
 A:Accession: J01434
 A:Molecule type: DNA
 A:Residues: 1-404 <HUL>
 A:Cross-references: EMBL:X58140; NID:922937; PIDN:CAA41148.1; PID:922938
 A:Note: the authors translated the codon AAA for residue 358 as Ala
 C:Comment: Unlike Epstein-Barr nuclear antigen 1 (EBNA-1) (see PIR:S42440), which is exp
 y of nasopharyngeal carcinoma tumor biopsies.
 C:Genetics:
 A:Gene: LMP1; BNLF-1
 A:Introns: 90/1; 119/1
 C:Superfamily: Epstein-Barr virus latent membrane protein
 C:Keywords: oncogene; tandem repeat; transmembrane protein
 F:25-45/Domain: transmembrane #status predicted <TM1>
 F:51-71/Domain: transmembrane #status predicted <TM2>
 F:76-97/Domain: transmembrane #status predicted <TM3>
 F:104-124/Domain: transmembrane #status predicted <TM4>
 F:140-160/Domain: transmembrane #status predicted <TM5>
 F:166-186/Domain: transmembrane #status predicted <TM6>
 F:236-332/Region: 11-residue repeats (P-D-N-T-D-D-N-G-P-Q-D)

Query Match 10.7%; Score 100.5; DB 1; Length 404;
 Best Local Similarity 23.6%; Pred. No. 0.27; Mismatches 53; Indels 55; Gaps 8;
 Matches 42; Conservative 28; Mismatches 53; Indels 55; Gaps 8;

OY 16 LSFSPPLTMIALISMSSWFNIMNALSOLGHAV-----KSSV 53
 Db 21 LSSSIGLALLLLALLFLWLYI---VMSDWTGALLVYSPALMIIILIIIFRRDL 77
 OY 54 APIPLGLAIGLIVYGLRNLYMSRVKGSLLISMVGLNLIGVEVYG-WIHF-- 110
 Db 78 CPLGLGLLLMLTITLLALNMLHQAALYIGIVLEFEGCL--VGVIIWYFLEI 129
 OY 111 -----VSVPFLSIIAFVIAISILDKSWIANV-----LIIGHIAM-YLH 149
 Db 130 LMRIGATIMQLAFLAFLAIIILITIALYLOQNMWTLVLDMLMLLIMWYTH 187

RESULT 11

E82442

probable peptide ABC transporter, permease protein VCA0590 [imported] - Vibrio cholerae
 C:Species: Vibrio cholerae
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 17-Nov-2000

C:Accession: E82442
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.
 chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers,
 J.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen vibrio cholerae.
 A:Reference number: A82035; MUID:20406833
 A:Accession: E82442
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-360 <HEI>
 A:Cross-references: GB:AE004389; GB:AE003853; NID:99657989; PIDN:AAF96492.1; GSPDB:GN
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor
 C:Genetics:
 A:Gene: VCA0590
 A:Map position: 2
 C:Superfamily: oligopeptide permease protein oppB

Query Match 10.5%; Score 98.5; DB 2; Length 360;
 Best Local Similarity 25.8%; Pred. No. 0.36; Mismatches 52; Indels 17; Gaps 6;
 Matches 34; Conservative 29; Mismatches 52; Indels 17; Gaps 6;

OY 1 MSLKHSWMDMIFILSFSPPLTMIALISMSSWFNIMNALSOLGHAVSSVAPIFNLG 60
 Db 131 VSISGLWSTLLIYLI--SIPLG-ISKAIHHSRFDVSSAVIIGYAIIP-----G 178
 OY 61 LAIGGILVIVGLRNLYMSRVKGSLLISMVGLNLIGVEVYGWIHFVSVFLSLII 120
 Db 179 FLPAIILILIFASGNYGSWFPLRGVSDNFASLPWYQOVD--YFW-HLTPTL--AMVI 233
 OY 121 AYPFAISILDKS 132
 Db 234 GGFATLSMLTKN 245

RESULT 12

A71694

hypothetical protein RP368 - Rickettsia prowazekii

C:Species: Rickettsia prowazekii
 C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
 C:Accession: A71694
 R:Andersson, S.G.E.; Zomrodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark
 Nature 396, 133-140, 1998
 A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
 A:Reference number: A71690; MUID:99039499
 A:Accession: A71694
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-280 <AND>
 A:Cross-references: GB:AJ235271; GB:AJ235269; NID:93868717; PIDN:CAA14827.1; PID:9386
 A:Experimental source: strain Madrid E
 C:Genetics:
 A:Gene: RP368

Query Match 10.4%; Score 97.5; DB 2; Length 280;
 Best Local Similarity 27.7%; Pred. No. 0.34; Mismatches 50; Indels 23; Gaps 6;
 Matches 39; Conservative 29; Mismatches 50; Indels 23; Gaps 6;

OY 3 LKHSWMDMIFILSFSPPLTMIALISMSSW---FNIMNALSOLGHAVSSVAPIFN- 58
 Db 119 LDLDNMVLPILINSFIIIVILILKGMGLFLRAFGRKDLIDLG-----KPAELYRM 173
 OY 59 LGIAIGLIVYGLRNLYMSRVKGSLLISMVGLNLIGVEVYGWIHFVSV----- 113
 Db 174 LGLSISNGLAALVTGLS---AQNNGFADINMGYGVALVIGIGAITIGQIFLNINMENA 229
 OY 114 ---LF--FLSIYAFIAISIL 129
 Db 230 LKEIFACTIGILFYFISLIL 250

RESULT 13
C82445
hypothetical protein VCA0550 [imported] - *Vibrio cholerae* (group O1 strain N16961)
C:Species: *Vibrio cholerae*
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
C:Accession: C82445
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F. L., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
A:Reference number: A82035; MUID:20406833
A:Accession: C82445
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-360 <HEI>
A:Cross-references: GB:AE004386; GB:AE003853; NID:99657957; PIDN:AAF96452.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VCA0550
A:Map position: 2

Query Match 10.4%; Score 97.5; DB 2; Length 360;
Best Local Similarity 25.9%; Pred. No. 0.43;
Matches 45; Conservative 29; Mismatches 49; Indels 51; Gaps 10;

OY 8 WMDITFILS--FSEPLMTALAI-----SMSSW-----FNTMNNALS-----DLGHAV 49
Db 197 WODVVISIMGCFEFLSLVLGVYWTDSPEFMAVFNFLPVLPLDGGHIL 256
OY 50 KSSVAPFNGLAIGLIVGLVGRNLYSMRKGSLISMGV-----FLNLIGFDEVYVG 105
Db 257 KS---VFSSNMKSKGVLICVAILGGI-----ALSYGLTLFGLLGMGLDIVE 305
OY 106 W-----IHFL-----VSYLEPLSIATFIAT-----SLDKSWIAVLIIIG 141
Db 306 WRORHSHLLPLNRVQMWSTIWFYFALVSGLIATIIIGFASGDTLLSLPILIG 359

RESULT 14
S26026
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 6 - *Caenorhabditis elegans* mitochondrion
C:Species: *mitochondrion Caenorhabditis elegans*
C>Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 20-Jun-2000
C:Accession: S26026; S25799
R:Okimoto, R.; Macfarlane, J.L.; Clary, D.O.; Wolstenholme, D.R.
Genetics 130, 471-498, 1992
A:Title: The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris*
A:Reference number: S26014; MUID:92201635
A:Accession: S26026
A:Molecule type: DNA
A:Residues: 1-144 <OKI>
A:Cross-references: EMBL:X54252; NID:913988; PID:92654243
R:Okimoto, R.; Macfarlane, J.L.; Wolstenholme, D.R.
Nucleic Acids Res. 18, 6113-6118, 1990
A:Title: Evidence for the frequent use of TGG as the translation initiation codon of mit
A:Reference number: S13139; MUID:91045077
A:Accession: S25799
A:Molecule type: DNA
A:Residues: 1-25 <OKX>
A:Cross-references: EMBL:X54252
C:Genetics:
A:Gene: ND6
A:Genome: mitochondrion
A:Genetic code: SGC4
A:Start codon: ATA
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 6
C:Keywords: mitochondrion; NAD; oxidoreductase

Query Match 10.3%; Score 97; DB 2; Length 144;
Best Local Similarity 23.5%; Pred. No. 0.21;

Matches 31; Conservative 23; Mismatches 40; Indels 38; Gaps 4;

OY 14 FILSEFPLMTALAISSMSWFNTMNNALS--DLGHAVSSVAPFNGLAIGLIVYGL 73
Db 27 FFLIFSLFSPVISMNHIF-----YFICLFLSGILVVIIVF 67
OY 74 RNLVSMRKGSL-----IISMGVFLNLIGFDEVYGMIF--LVSVLF 115
Db 68 SLSKINNVASVMAVFLLLSLMLVFSPVLYSSYLGSGFYISY-WFICFPLVLCLEF 126
OY 116 FLSTIAFIAIS 127
Db 127 FMNFSYFLNFS 138

RESULT 15
H82822
NADH-ubiquinone oxidoreductase, N0014 subunit XF0318 [imported] - *Xylella fastidiosa*
C:Species: *Xylella fastidiosa*
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: H82822
R:Anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: H82822
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-485 <SLIM>
A:Cross-references: GB:AE003884; GB:AE003849; NID:99105127; PIDN:AAF83129.1; GSPDB:GN
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvares, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier
as-Neto, E.; Docena, C.; El-Dorty, H.; Facincan, A.P.; Ferreira, A.J.S.
submitted to Genbank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Fr
J.D.; Jungueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
Rodrigues, V.; Rosa, A.J. de M.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri,
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF0318
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 2

Query Match 10.2%; Score 96.5; DB 2; Length 485;
Best Local Similarity 25.8%; Pred. No. 0.68;
Matches 41; Conservative 22; Mismatches 59; Indels 37; Gaps 6;

OY 15 ILTSPFLMTALAISSMSWFNTMNNALS--DLGHAVSSVAPFNGLAIGLIVYGL 72
Db 22 LIGSTFALLMLDFTGQAR--KWTHTLSVAILGVLSMLVAGCGGCAARHGMF----- 74
OY 73 LRNLVSMRKGSLISMGVFLNLIGFDEVYGMIFLVSFL--FLSTIAFIAISILD 130
Db 75 -----VRDSAADVWKVIVLVLSALSYGMSYLRERLFGCEIPVLVLFATLGM- 124
OY 131 KSWIAVLLIIGHIAMWYLFHASEIPRCAIPELLAVSF 169
Db 125 -----VIVSAGHLMYVIGL-----ELLALCSY 147

Search completed: May 30, 2001, 15:24:21
Job time: 91 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2001, 15:24:05 ; Search time 9.55 Seconds
(without alignments)
660.001 Million cell updates/sec

Title: US-09-382-242-33

Perfect score: 942
Sequence: 1 MSLNKSWMDMIFILSF...AVSFLLPYIRDFKSYTKR 184

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	100.5	10.7	404	1	LMP1_EBYC
2	97.5	10.4	280	1	P29362 epstein-bar
3	97	10.3	144	1	P24851 rickettsia
4	96	10.2	419	1	P24885 caenorhabdi
5	94.5	10.0	397	1	007523 bacillus su
6	94.5	10.0	1228	1	062668 rattus norv
7	94	10.0	394	1	098993 arabidopsis
8	93.5	9.9	460	1	P41733 saccharomyc
9	93	9.9	383	1	092k47 helicobacte
10	90.5	9.6	532	1	P42946 saccharomyc
11	89.5	9.5	461	1	P37555 bacillus su
12	89	9.4	226	1	025551 helicobacte
13	89	9.4	243	1	000275 apis mellif
14	89	9.4	277	1	P75314 mycoplasma
15	88.5	9.4	397	1	057685 methanococ
16	88	9.3	313	1	015849 homo sapien
17	87.5	9.3	387	1	099824 rhipicephal
18	87.5	9.3	453	1	019025 mactrosceid
19	87.5	9.3	455	1	P19338 rattus norv
20	87.5	9.3	475	1	P30545 mus musculu
21	87.5	9.3	1202	1	053092 lactobacilli
22	87	9.2	261	1	031183 arabidopsis
23	87	9.2	261	1	P55777 gadus morhu
24	87	9.2	551	1	P34202 erwinia car
25	86.5	9.2	198	1	P44472 haemophilus
26	86.5	9.2	305	1	P18015 clostridium
27	86.5	9.2	364	1	032256 bacillus su
28	86	9.1	275	1	P33914 escherichia
29	86	9.1	299	1	P42953 bacillus su
30	86	9.1	319	1	057882 methanococ
31	86	9.1	341	1	P75262 mycoplasma
32	86	9.1	369	1	P38893 diosiphila
33	86	9.1	622	1	037381 acanthamoeb
					P24010 bacillus su

34	85.5	9.1	492	1	NU4M_CHOCR	P48915 chondrus cr
35	85	9.0	279	1	DMSC_HAEIN	P45002 haemophilus
36	85	9.0	393	1	YHPC_ECOLI	P21229 escherichia
37	85	9.0	1029	1	YEC5_YEAST	P43571 saccharomyc
38	84.5	9.0	230	1	ATP6_ASTPE	033823 asterina pe
39	84.5	9.0	397	1	UT2_RABIT	028614 oryctolagus
40	84.5	9.0	430	1	REEB_SALT	P26400 salmonella
41	84.5	9.0	450	1	A2AB_HUMAN	P18089 homo sapien
42	84.5	9.0	474	1	ST24_SCHPO	010071 schizosacch
43	84.5	9.0	660	1	P390_MYCPN	P75207 mycoplasma
44	84.5	9.0	1203	1	ALAB_ARATH	098455 arabidopsis
45	84	8.9	332	1	YPHD_ECOLI	P77315 escherichia

ALIGNMENTS

RESULT	1	STANDARD	PRT	404 AA
LMP1_EBYC				
ID	LMP1_EBYC			
AC	P29362			
DT	01-DEC-1992 (Rel. 24, Created)			
DT	01-DEC-1992 (Rel. 24, Last sequence update)			
DT	01-OCT-1994 (Rel. 30, Last annotation update)			
DE	LATENT MEMBRANE PROTEIN 1 (LMP-1) (P63).			
GN	BNLF1.			
OS	Epstein-Barr virus (strain Cao) (Human herpesvirus 4).			
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;			
OC	Gammaherpesvirinae; Lymphocryptovirus.			
OX	NCBI_TaxID=31525;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92013956; PubMed=1681026;			
RA	Hu L.F., Zaborovsky E.R., Chen F., Cao S.L., Ernberg I., Klein G.,			
RA	Winberg G.;			
RT	"Isolation and sequencing of the Epstein-Barr virus BNLF-1 gene			
RT	(LMP1) from a Chinese nasopharyngeal carcinoma.";			
RL	J. Gen. Virol. 72:2399-2409(1991).			
CC	1- FUNCTION: THE LATENT MEMBRANE PROTEIN HAS TRANSFORMING ACTIVITY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: X58140; CAA41148.1; -			
DR	EMBL: D10059; BAA00948.1; -			
DR	PIR: J01434; LABECA.			
DR	PIR: S21660; S21660.			
KW	Transmembrane; Phosphorylation; Transforming protein.			
FT	DOMAIN 1 24			
FT	TRANSMEM 25 44			
FT	TRANSMEM 52 72			
FT	TRANSMEM 77 97			
FT	TRANSMEM 105 125			
FT	TRANSMEM 139 159			
FT	TRANSMEM 166 186			
FT	DOMAIN 187 404			
FT	SEQUENCE 404 AA; 43769 MW; D04536D3B5F82E CRC64;			

Query Match 10.7%; Score 100.5; DB 1; Length 404;
Best Local Similarity 23.6%; Pred. No. 1.1;
Matches 42; Conservative 28; Mismatches 53; Indels 55; Gaps 8;

QY 16 LSFSPFLMIALAISMSSEFNIMNNAISD-LGHAV-----KSSV 53
DB 21 LSSSGALALLALLALFLWLYT---VMSDWGALVLYSPALMKITIIILIFRRRL 77
QY 54 APIFNLAIGIILVIVGLRNLYSWSRVKCSLIISMGVFLNLGVDEVYG-WIHFL- 110

```

Db      78 CPLGGILLMLITLLTLLALNMLHGOALYGLVIFEGCLL-----VVGIMYIFLEI 129
Oy      111 -----VSLEFLSIAYFAINISILDKSMIV-----LLIIGHIAMV-YLH 149
Db      130 LMRGATIMOLLAFLLAFLLAIIILITLALYLOQNMWTLVLDLMLLPMALIMYIH 187

RESULT 2
Y368_RICPR STANDARD: PRT: 280 AA.
ID Y368_RICPR
AC Q92DGL;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE HYPOTHETICAL PROTEIN RP368.
GN RP368.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E.;
RX MEDLINE=98039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Slikeritz-Poten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria."
RL Nature 396:133-140(1998).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ235271; CAA14827.1; -
DR Hypothetical protein; Transmembrane.
FT TRANSMEM 3 23
FT TRANSMEM 52 72
FT TRANSMEM 81 101
FT TRANSMEM 123 143
FT TRANSMEM 196 216
FT TRANSMEM 233 253
FT TRANSMEM 280 AA; 30554 MW; CEF294794D843FPA1 CRC64;
SQ SEQUENCE

Query Match 10.4%; Score 97.5; DB 1; Length 280;
Best Local Similarity 27.7%; Pred. No. 1.3;
Matches 39; Conservative 29; Mismatches 50; Indels 23; Gaps 6;
Oy 3 LNRKSHMDMIFILSFSPPLTMIALALISMSSW---FNIMNNALSDLGHAIVSSVAPIRN- 58
Db 119 LDLDMLVPLILINSFIVIVILILKGNLGLFRAFGNKDLIDLG-----KPAELVRM 173
Oy 59 LGLAIGGLIIVYGLRNLYKSRVKGSLISMGVFLNLIGVFDEYGGIHLVSV----- 113
Db 174 LGLISNGLALVTGLS---AQNVGADINMGVGVALVIGAILIIGQIFLNINNFNA 229
Oy 114 ---LF--FLSIAYFAISIL 129
Db 230 LKEIFACITGILFYFISIL 250

RESULT 3
NU6M_CAEBL STANDARD: PRT: 144 AA.
ID NU6M_CAEBL
AC P24885;

```

```

DT 01-MAR-1992 (Rel. 21, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 6 (EC 1.6.5.3).
GN NDB.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=92201635; PubMed=1551572;
RA Okimoto R., Macfarlane J.L., Clary D.O., Wolstenholme D.R.;
RT "The mitochondrial genomes of two nematodes, Caenorhabditis elegans
RT and Ascaris suum."
RL Genetics 130:471-496(1992).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X54252; CAA38153.1; -
DR PIR: S26026; S26026.
DR OXidoreductase; NAD; Ubiquinone; Mitochondrion.
KW SEQUENCE 144 AA; 16788 MW; 04F51D0E354B40F9 CRC64;
SQ SEQUENCE

Query Match 10.3%; Score 97; DB 1; Length 144;
Best Local Similarity 23.5%; Pred. No. 0.81;
Matches 31; Conservative 23; Mismatches 40; Indels 38; Gaps 4;
Oy 14 FILSPFLTMTALISMSSWFNIMNNALSDLGHAIVSSVAPIRNGLIIVYGL 73
Db 27 FFLISLFLSPVLSMHSIMHS-----YFCLLFLSGIFVILVYF 67
Oy 74 RNLYKSRVKGSL-----IISMGVFLNLIGVFDEYGGIHLVSVLF 115
Db 68 SLSKINIVKSKMAVFLDLISMLYFSPVLYSYSLGSGFYYSY-WFICFLVCLLF 126
Oy 116 FLSTIAYFAIS 127
Db 127 FWNSSYFLNFS 138

RESULT 4
YHAP_BACSU STANDARD: PRT: 419 AA.
ID YHAP_BACSU
AC O07523;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 45.4 KDA PROTEIN IN SSPB-PRSA INTERGENIC REGION.
GN YHAP.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Noback M.A., Terpstra P., Holsappel S., Venema G., Bron S.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO M.JANNA SCHI MJ1024.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC Between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: Y14078; CAAT74423.1; -
 DR EMBL: 299109; CAB12830.1; -
 DR Subtilisin; BGI2992; ynap.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 16 36 POTENTIAL.
 FT TRANSMEM 186 206 POTENTIAL.
 FT TRANSMEM 235 255 POTENTIAL.
 FT TRANSMEM 283 303 POTENTIAL.
 FT TRANSMEM 318 338 POTENTIAL.
 FT TRANSMEM 340 360 POTENTIAL.
 FT TRANSMEM 369 389 POTENTIAL.
 SQ SEQUENCE 419 AA; 45429 MW; E6434F27E5E08A2 CRC64;

Query Match 10.28; Score 96; DB 1; Length 419;
 Best Local Similarity 25.78; Pred. No. 2.3;
 Matches 46; Conservative 35; Mismatches 68; Indels 30; Gaps 9;

QY 11 MIIFILSFPLFMIALAISMSWFMIMNNALSDIGHAVKSSVAPIN-----IGLAIGG 65
 Db 189 IMLEVITFS-VIMYASMIAM-----EVATKSSRWMEILISSMPIDOMRAKLIGIVG 242

QY 66 I--LIVIVGLRNLVSMRSRVKSLISMGVPLNT--IGVFDEVYGMIFLVSLEPFLSIIA 121
 Db 243 ITQALITIGAGSLSLKLNQKSETASVGAFLNTDVSATVIVAVIFLAVFYATLAA 302

QY 122 YF--IAISILDKSW-----IAVLIIIGH-IAMMYLHFASEIPRGAIPPELLAVSFPLPY 173
 Db 303 FLGSVSRIEDVQVITIPMTLLVYAGFMIAFMGLN-----APDAGFITVTSFIPFF 353

RESULT 5
 UT2_RAT STANDARD; PRT; 397 AA.

AC 062668;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE UREA TRANSPORTER, KIDNEY.
 GN SLCL14A2 OR UT2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Kidney;
 RX MEDLINE=95386714; PubMed=7657826;
 RA Smith C.P., Lee W.-S., Martial S., Knepper M.A., You G., Sands J.M.,
 Hediger M.A.;
 RT "Cloning and regulation of expression of the rat kidney urea
 RT transporter (UT2).";
 RT J. Clin. Invest. 96:1556-1563(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Kidney inner medulla;
 RX MEDLINE=97117152; PubMed=8958221;
 RA Shiyakul C., Steel A., Hediger M.A.;
 RT "Molecular cloning and characterization of the vasopressin-regulated
 RT urea transporter of rat kidney collecting ducts";
 RT J. Clin. Invest. 98:2580-2587(1996).
 CC -1- FUNCTION: SPECIALIZED LOW-AFFINITY UREA TRANSPORTER. MEDIATES UREA
 CC TRANSPORT IN KIDNEY.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: KIDNEY.
 CC -1- SIMILARITY: TO OTHER MAMMALIAN UREA TRANSPORTERS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U09957; AAA84392.1; -
 DR EMBL: U77971; AAB50197.1; ALT_INIT.
 KW Transport; Transmembrane; Glycoprotein.
 FT TRANSMEM 68 85 POTENTIAL.
 FT TRANSMEM 92 109 POTENTIAL.
 FT TRANSMEM 115 135 POTENTIAL.
 FT TRANSMEM 143 163 POTENTIAL.
 FT TRANSMEM 172 192 POTENTIAL.
 FT TRANSMEM 239 257 POTENTIAL.
 FT TRANSMEM 264 280 POTENTIAL.
 FT TRANSMEM 287 303 POTENTIAL.
 FT TRANSMEM 309 329 POTENTIAL.
 FT TRANSMEM 331 351 POTENTIAL.
 FT CARBOHYD 210 210
 SQ SEQUENCE 397 AA; 43237 MW; EA2068EA25661224 CRC64;

Query Match 10.08; Score 94.5; DB 1; Length 397;
 Best Local Similarity 22.28; Pred. No. 2.8;
 Matches 50; Conservative 30; Mismatches 60; Indels 85; Gaps 12;

QY 4 NKHSMDMIIFILSFPLFMIALAISMSWFMIMNNALSDIGHAVKSSVAPINFLAI 63
 Db 140 NYWMLLPVIVSMTCPIILSSALSTVFSK-----DL-----PVTLPLNI 181

QY 64 GGLI-IVIVGLRNLV-----SWSRVKSLI--ISMGVFLNLIGVFDEVY 104
 Db 182 AVTLVLAATGHYNLFEPFKLLQPAVTPNTITWSDVQVPLLRAPVGI-----GVY 233

QY 105 G-----VHFLVSLEPFL-----SIIAFIAISILD-----KSMIV 136
 Db 234 GCDNPWTGGIFLVALFVSSPLICLAAIGSTIGMALAISTATPPDSIVFGICGFNSTLAC 293

QY 137 LLING--HIAMMYLH-----FASEIPRGAIPPELLAVSFPLP 171
 Db 294 IATGMEYVITWOTHLAIACALFRAVI--GAALANMLSVGLP 336

RESULT 6
 ALA5_ARATH STANDARD; PRT; 1228 AA.

AC 09563;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE POTENTIAL PHOSPHOLIPID-TRANSPORTING ATPASE 5 (EC 3.6.3.13).
 GN ALA5 OR ATIG27200 OR F28P22.11.
 OS Arabidopsis thaliana (Mouse ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hutzar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurts D.B., Kwan A., Lam B.,

AC 092KW7;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE VIRULENCE FACTOR MVIN HOMOLOG.
 GN MVIN OR JHP0817.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 OC NCBI_TaxID=85963;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=99120557; PubMed=9923682;
 RX Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonghe B.L., Carmel G.,
 RA Tummino P.J., Caruso A., Ulla-Nickelsen H., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen Helicobacter pylori";
 RL Nature 397:176-180(1999).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE MVIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE001511; AAD06398.1; -
 CC Transmembrane.
 KW Transmembrane.
 FT TRANSMEM 4 24 POTENTIAL.
 FT TRANSMEM 50 70 POTENTIAL.
 FT TRANSMEM 95 115 POTENTIAL.
 FT TRANSMEM 122 142 POTENTIAL.
 FT TRANSMEM 155 175 POTENTIAL.
 FT TRANSMEM 228 248 POTENTIAL.
 FT TRANSMEM 258 278 POTENTIAL.
 FT TRANSMEM 292 312 POTENTIAL.
 FT TRANSMEM 336 356 POTENTIAL.
 FT TRANSMEM 366 386 POTENTIAL.
 FT TRANSMEM 387 407 POTENTIAL.
 FT TRANSMEM 428 448 POTENTIAL.
 FT TRANSMEM 460 AA; 51484 MW; 0E957BB6023C1FC1 CRC64;
 SQ SEQUENCE

Query Match
 Best Local Similarity 9.9%; Score 93.5; DB 1; Length 460;
 Matches 31; Conservative 33; Mismatches 55; Indels 19; Gaps 6;

OY 15 ILSEFPLTMA-LAISMSWFNINNALSDLGHAVKSSVAPIFNLGILIVTGL 73
 DB 332 ITSQVEFLYLGLLGLFGLTFLKFLSLMLYAKLEQKRAKISLSLFL-LGLAASLSIMPLIGV 390
 OY 74 RNLSMNRVSGSLISGV-----PLNLGVFDEYVGIMHPLVSLFLSLIAVIAISIL 129
 DB 391 LGLALANLSLGLFLFVLTIAFGFQSLGIITIKLSWL-----VILPLACVEILLALAF- 444
 OY 130 DKSMIAVLLIIGHIIMY 147
 DB 445 -KSWT-----HLXLYF 455

RESULT 9
 YJRK YEAST STANDARD; PRT; 383 AA.
 AC P42946;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-DEC-1998 (Rel. 37, Last annotation update)

DE HYPOTHETICAL 41.5 KDA PROTEIN IN GZF3-IME2 INTERGENIC REGION.
 GN YJL108C OR J0811.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C / FY1679;
 RC MEDLINE=96090136; PubMed=7483851;
 RX Rasmussen S.W.;
 RA "A 37.5 kb region of yeast chromosome X includes the SMI1, MEE2, GSH1
 RT and GSD3 genes, a TCP-1-related gene, an open reading frame similar
 RT to the Dnl80 gene, and a tRNA(Arg).";
 RL Yeast 11:873-883(1995).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: TO S.POMBE SPAC16A10.01.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X85021; CAAS9386.1; -
 CC DR EMBL: Z49383; CA89402.1; -
 CC DR SGD: S0003644; YJL108C.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 66 86 POTENTIAL.
 FT TRANSMEM 88 108 POTENTIAL.
 FT TRANSMEM 118 138 POTENTIAL.
 FT TRANSMEM 142 162 POTENTIAL.
 FT TRANSMEM 181 201 POTENTIAL.
 FT TRANSMEM 217 237 POTENTIAL.
 FT TRANSMEM 242 262 POTENTIAL.
 FT TRANSMEM 272 292 POTENTIAL.
 FT TRANSMEM 294 314 POTENTIAL.
 FT TRANSMEM 353 373 POTENTIAL.
 SQ SEQUENCE 383 AA; 41488 MW; 2030F73D1823CE04 CRC64;

Query Match
 Best Local Similarity 9.9%; Score 93; DB 1; Length 383;
 Matches 51; Conservative 40; Mismatches 59; Indels 86; Gaps 12;

OY 8 WMDMIFLSEFPLTMA-LAISMSWFNINNALSDLGHAVKSSVAPIFNLGILIGI- 66
 DB 88 WVNLA--SFMGLCYGSLQFLLSQKSYMSVNF-----ISASIVVF-CGRAPGISP 138
 OY 67 -----LIVYGLRNLYSGSRVG-----SLISMGVPLNL 96
 DB 139 RSHICFAYVQSGSLALPLGYILLGALDELQSRSLVAGVAFYAIYSLFLGIGITLG- 197
 OY 97 IGVEDEVYGVHIF-----LVSYLF-PLSTIAYFAISIDKSMIAVLLI----- 140
 DB 198 -----SALFGMNYHATNEISCPLISPFRLFPAPATISISLNGAHSIDLPVWFLSC 253
 OY 141 -GHIMMYL--HFA-----SEIPRGAIPPELL-AVSEFLP 171
 DB 254 TGYVVTWAGHKFANSTETFAALAFYIGVNLYSRIWKGLAVSAMLPATFYGVV 309

RESULT 10
 YABM_BACSU STANDARD; PRT; 532 AA.
 AC P37555;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE HYPOTHETICAL 57.4 KDA PROTEIN IN MFD-DIVIC INTERGENIC REGION.
 GN YABM.

OS Bacillus subtilis.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 CC NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=96051385; PubMed=7584024;
 RA Ogasawara N., Nakai S., Yoshikawa H.;
 RT "Systematic sequencing of the 180 kilobase region of the Bacillus
 RT subtilis chromosome containing the replication origin."
 RL DNA Res. 1:1-14(1994).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
 CC -1- SIMILARITY: STRONG. TO B. SUBTILIS SPOVA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 DR EMBL: D26185; BAA05292.1; -
 DR EMBL: Z99104; CAB1833.1; -
 DR Subtilist: BG10120; yabm.
 DR InterPro: IPR002797; -
 DR Pfam: PF01943; Polysacc_synt. 1.
 KM Hypothetical protein; Transmembrane.
 FT TRANSMEM 25 45 POTENTIAL.
 FT TRANSMEM 65 85 POTENTIAL.
 FT TRANSMEM 109 129 POTENTIAL.
 FT TRANSMEM 134 154 POTENTIAL.
 FT TRANSMEM 179 199 POTENTIAL.
 FT TRANSMEM 203 223 POTENTIAL.
 FT TRANSMEM 248 268 POTENTIAL.
 FT TRANSMEM 302 322 POTENTIAL.
 FT TRANSMEM 344 364 POTENTIAL.
 FT TRANSMEM 371 391 POTENTIAL.
 FT TRANSMEM 392 412 POTENTIAL.
 FT TRANSMEM 425 445 POTENTIAL.
 FT TRANSMEM 459 479 POTENTIAL.
 FT TRANSMEM 494 514 POTENTIAL.
 SQ SEQUENCE 532 AA; 57357 MW; DAAACAF32903E553 CRC64;
 Query Match 9.6%; Score 90.5; DB 1; Length 532;
 Best Local Similarity 24.7%; Pred. No. 6.7; Mismatches 84; Indels 43; Gaps 10;
 Matches 53; Conservative 35;
 QY 5 KHSNMDMIFILSFPLTMAIALISMSSWFNINNALSDLGHAVKSSVAPFENLGLAI- 63
 DB 8 KRHWIWCAGFLLAGVITKLSAVRVPPQNI---VGDVGFIYQOYVPF--LGIAVM 61
 QY 64 ---GGIIVVGLNLSKMSRWKSLIISMGVFLNLIGV--FDEVY-----GWIH 108
 DB 62 LSTSGFVIIISKLMNDYSEKNNHTLIKIS-ALFSLIGILLFCLLYGAVPIALFMDSH 120
 QY 109 F--LVSLFPLTIAFYI-----AISILDKSWIAVLLIIGHIAMVTL-H 149
 DB 121 LAVLIQVAATAFLLFPVALLRGFGCRHEMLPSALSMTEQFLRVAVLGL-LSFVLVKK 179
 QY 150 FASEIPRGAAPPELLAVSEFLPFYIRDFYSYTKR 184
 DB 180 GASLYTAGAANAAGSLAGSLVALILLGFFWFKTKR 214
 RESULT 11
 MYIN_HELPY STANDARD; PRT; 461 AA.
 AC 025551;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE VIRulence factor MYIN HOMOLOG.
 GN MYIN OR HP0885.
 OS Helicobacter pylori (Campylobacter pylori).
 CC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 CC Helicobacter.
 CC NCBI_TaxID=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26695 / ATCC 700392;
 RX MEDLINE=97394467; PubMed=9252185;
 RA Tomb J.-F., White O., Keriavage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 RA McInerney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Meldrum J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.;
 RT "The complete genome sequence of the gastric pathogen Helicobacter
 RT pylori."
 RL Nature 388:539-547(1997).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE MYIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 DR EMBL: AE000598; AAD07933.1; -
 DR TIGR: HP0885; -
 KM Transmembrane.
 FT TRANSMEM 5 25 POTENTIAL.
 FT TRANSMEM 51 71 POTENTIAL.
 FT TRANSMEM 96 116 POTENTIAL.
 FT TRANSMEM 123 143 POTENTIAL.
 FT TRANSMEM 156 176 POTENTIAL.
 FT TRANSMEM 229 249 POTENTIAL.
 FT TRANSMEM 259 279 POTENTIAL.
 FT TRANSMEM 294 314 POTENTIAL.
 FT TRANSMEM 337 357 POTENTIAL.
 FT TRANSMEM 372 392 POTENTIAL.
 FT TRANSMEM 402 422 POTENTIAL.
 FT TRANSMEM 429 449 POTENTIAL.
 SQ SEQUENCE 461 AA; 51618 MW; 62B7DAA43967A46C CRC64;
 Query Match 9.5%; Score 89.5; DB 1; Length 461;
 Best Local Similarity 22.5%; Pred. No. 7.1; Mismatches 56; Indels 19; Gaps 6;
 Matches 31; Conservative 32;
 QY 15 ILSPFPLTMAIALISMSSWFNINNALSDLGHAVSSVAPFENLGLIIVVGL 73
 DB 333 ITSGVFSLYLGLPRLTFLFSLMLYAKLEOKRAAKISLSIF-LGLASLSIMPLIGV 391
 QY 74 RNLYSWSRWKSGSLIISMGV--FLNLIGVFDEVYGGIHEFLVSLFSLIYAFIAISIL 129
 DB 392 LGLALANSLSGLFLFVLTIAKAGFGLFGLIKNLKSWL-----VILFLACVELLLLAFL- 445
 QY 130 DKSWIAVLLIIGHIAMV 147
 DB 446 -KSWVT-----HLIYLFY 456
 RESULT 12
 ATP6_APIIT STANDARD; PRT; 226 AA.
 AC 000275;
 DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ATP SYNTHASE A CHAIN (EC 3.6.1.34) (PROTEIN 6).
GN ATP6.
OS Apis mellifera ligustica (Common honeybee).
OG Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
OC Apoidea; Apidae; Apis.
OX NCBI_TaxID=7469;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thorax;
RX MEDLINE=92261310; PubMed=1533894;
RA Crozier R.H., Crozier Y.C.;
RT "The cytochrome b and ATPase genes of honeybee mitochondrial DNA";
RL Mol. Biol. Evol. 9:474-482(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Thorax;
RX MEDLINE=93114603; PubMed=8417993;
RA Crozier R.H., Crozier Y.C.;
RT "The mitochondrial genome of the honeybee Apis mellifera: complete
sequence and genome organization.";
RL Genetics 133:97-117(1993).
CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE.
CC -1- SUBUNIT F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), Epsilon(1). CF(0)
HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L06178; AAB96802.1; -
DR EMBL; M87065; AAA31634.1; -
DR InterPro; IPR000568; -
DR Pfam; PF00119; ATP_syn_A; 1.
DR PRINTS; PR00123; ATPASE_A.
DR PROSITE; PS00449; ATPASE_A; 1.
KW Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
SQ SEQUENCE 226 AA; 26694 MW; B6F78A57B06A1D4 CRC64;

Query Match 9.4%; Score 89; DB 1; Length 226;
Best Local Similarity 24.0%; Pred. No. 4.3;
Matches 37; Conservative 42; Mismatches 51; Indels 24; Gaps 9;

QY 1 MSLNKSHNMIMFILSFSPPLT---MIALAISMSMFPN-----IMNNAISDLCHAVKSS 52
DB 77 ISLMXYIMINISFLIPYVFTLSHLINMLSLTLMFSEFLIYINNYIMFSLHVLPLN 136
QY 53 VAPFENGLAIGILYIVGLRNLX--SWS---RVKGLISIMGVNPLIGVPEYVGM 107
DB 137 -SPVFLMN-----FMVITELISLITRPWTLSIRLSANL-ISGHLITLGLIF--ISNFI 186
QY 108 HFL-VSVLFLSIAYPFAISILDKSWIAVLLI 140
DB 187 SLIPINIMIONMLLTLEIFMSMIGSYVFSILLI 220

RESULT 13
Y23B_MYCPN STANDARD; PRT; 243 AA.

AC P75314;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCN-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MG323.1 HOMOLOG (P01_0RF243).
GN MPN469 OR MP372.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreich R., Hilbert H., Plagens H., Plank E., Li B.-C.,
Hermann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000035; AAB96020.1; -
DR Hypothetical protein; Transmembrane.
KW TRANSMEM 38 58 POTENTIAL.
FT TRANSMEM 99 119 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 204 224 POTENTIAL.
SQ SEQUENCE 243 AA; 28278 MW; DFD735C3DD876FE CRC64;

Query Match 9.4%; Score 89; DB 1; Length 243;
Best Local Similarity 24.7%; Pred. No. 4.5;
Matches 36; Conservative 23; Mismatches 41; Indels 46; Gaps 8;

QY 13 IFILSFSEPLTMIALAI-----SMSSMFINMNAISDLGHVKSVAPIFNL 59
DB 41 LFLISFVTAMFLVGLIFHTDPTLNDQNGISGFYLFN-----YAKP-ADIFNA 90
QY 60 GLAIG---GLIIVYGLRNLXSVKRVKGLIISMGVFLNLIGV--DEVYGMIFLVS 113
DB 91 NFVYSISSEGIALLALGLSLF-----LMIFLG-YRNAISLFIKSQTKMERVIFST 141
QY 114 LFLSIAYPFAISILDKSWIAVLLI 139
DB 142 GFYFSVVAYCF-----WIALML 159

RESULT 14
Y233_METUA STANDARD; PRT; 277 AA.

AC 057685;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN M0233.
GN M0233.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,

RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
RA Overbeek R., Kirkness E.F., Weissknob R.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Meldrum J.F., Ruhnmann J.L., Nguyen D.,
RA Uitterlbeck T.R., Kelley J.M., Peterson J.D., Sedow P.M., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii".
RT Science 273:1058-1073(1996).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO MAMMALIAN UREA TRANSPORTER.
CC -1- SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U67479; AAB98226.1; -
CC
CC DR TIGR: MJ0233; -
CC
CC KW Hypothetical protein; Transmembrane.
CC FT TRANSMEM 23 43
CC FT TRANSMEM 61 81 POTENTIAL.
CC FT TRANSMEM 117 137 POTENTIAL.
CC FT TRANSMEM 144 164 POTENTIAL.
CC FT TRANSMEM 197 217 POTENTIAL.
CC FT TRANSMEM 221 241 POTENTIAL.
CC FT TRANSMEM 243 263 POTENTIAL.
CC SQ SEQUENCE 277 AA; 31306 MW; 31F66D38BD552839 CRC64;

Query Match 9.4%; Score 89; DB 1; Length 277;
Best Local Similarity 22.3%; Pred. No. 5.1;

Matches 43; Conservative 27; Mismatches 53; Indels 70; Gaps 8;

QY 44 DLGAVKSSV---APFNLGLAIGIL-----IYVGLRLMYGMSRV-----GSSL 86
DB 53 DMIHINSFLYFIFILIGLIVSAIIIGYVNRIMKTVEELANAPDMNNTDILYKIL 112
QY 87 IISGVPNLIGVDEYVGMIFLVSFLSLIT-----AFYATISLDSMTAVL 137
DB 113 YIVGLVNLNIT-----FYFIPALFVGIFSLYISKLIGAFLLIISTL-----I 156
QY 138 LIIGHIIMWYHFASEIP-----RCALPELLAVSF----- 169
DB 157 FILSVISLMLYSKAEVNSYKGFYGFPERKEIPKMGIRITLIVITAIINFTISLIV 216
QY 170 LPEYIRDFYSY 182
DB 217 LPLNIIDIFISYS 229

RESULT 15

UT2_HUMAN STANDARD: PRT; 397 AA.
AC 015849;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE UREA TRANSPORTER, KIDNEY.
GN SLC14A2 OR UT2 OR HUT2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=66228053; PubMed=8647271;
RA Olives B., Sonia M., Mattei M.-G., Matassi G., Roussetlet G.,

RA Ripoché P., Cartton J.-P., Bailly P.;
RT "Molecular characterization of a new urea transporter in the human
RT kidney".
RL FEBS Lett. 386:156-160(1996).
CC -1- FUNCTION: SPECIALIZED LOW-AFFINITY UREA TRANSPORTER. MEDIATES UREA
CC TRANSPORT IN KIDNEY.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: KIDNEY.
CC -1- SIMILARITY: TO OTHER MAMMALIAN UREA TRANSPORTERS.
CC -1- SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X96969; CAA65657.1; -
CC
CC DR MIM: 601611; -
CC
CC KW Transport; Transmembrane; Glycoprotein.
CC FT TRANSMEM 68 85 POTENTIAL.
CC FT TRANSMEM 92 109 POTENTIAL.
CC FT TRANSMEM 115 135 POTENTIAL.
CC FT TRANSMEM 143 163 POTENTIAL.
CC FT TRANSMEM 172 192 POTENTIAL.
CC FT TRANSMEM 239 257 POTENTIAL.
CC FT TRANSMEM 264 280 POTENTIAL.
CC FT TRANSMEM 287 303 POTENTIAL.
CC FT TRANSMEM 309 329 POTENTIAL.
CC FT TRANSMEM 331 351 POTENTIAL.
CC FT CARBOHYD 210 210 N-LINKED (GLCNAC...) (POTENTIAL).
CC SQ SEQUENCE 397 AA; 43395 MW; 310AD64B1D9A9623 CRC64;

Query Match 9.4%; Score 88.5; DB 1; Length 397;
Best Local Similarity 21.7%; Pred. No. 7.3;

Matches 48; Conservative 29; Mismatches 59; Indels 85; Gaps 12;

QY 8 WMDNITIFLSFPLMTALAISSMFWNNALSDLGAVKSSVAPFIRNLGLAIGIL 67
DB 144 WLLPVIIMSMSCPIILSSALGTIFSKW-----PVFTLPENITVTL 185
QY 68 -IYVGLRNL-----SMSRYKGLI-----ISMGVFLNLIGVDEYV----- 105
DB 186 YLATGHNLFEPPTTLQAPASAMPNITWSEYQVPLLRALPVGI-----GOVYGCN 237
QY 106 -WIHLVSVLEFL-----STIAYFAISLID-----KSWIAYLLIT 140
DB 238 PWTGIFILALFISPLICLHAIGSTMGLMALTLATPFDSIYFGLCGFNSTLACIAG 297
QY 141 G-HIAMWYLH-----FASEIPRCALPELLAVSF 171
DB 298 GMFYVITWQTHLAIACALFPAAYL-GAALANMLSVGLPP 336

Search completed: May 30, 2001, 15:26:02
Job time: 117 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2001, 15:23:50 ; Search time 22.38 Seconds
(without alignments)
963.640 Million cell updates/sec

Title: US-09-382-242-33

Sequence: 1 MSLNKSHMMDMIFILSF.....AVSFPLPYIRDPKSYTKR 184

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_15:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.unclassified:*
13: sp.vertebrate:*
14: sp.virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	163.5	17.4	175	1	Q9V235	Q9V235 pyrococcus
2	144.5	15.3	156	1	059449	059449 pyrococcus
3	142	15.1	192	1	058000	058000 pyrococcus
4	129.5	13.7	183	1	059239	059239 pyrococcus
5	126.5	13.4	156	1	Q9V157	Q9V157 pyrococcus
6	118	12.5	165	1	Q9Y809	Q9Y809 aeropyrum p
7	109.5	11.6	512	2	Q9R0P4	Q9R0P4 burkholderi
8	105.5	11.2	401	2	Q9KS98	Q9KS98 vibrio chol
9	104.5	11.1	373	2	Q9RKO6	Q9RKO6 deinococcus
10	103.5	11.0	600	2	Q9KNN1	Q9KNN1 vibrio chol
11	102.5	10.9	293	5	Q9VGT0	Q9VGT0 dirosophila
12	101.5	10.8	365	8	Q9XMT1	Q9XMT1 tetrahymena
13	98.5	10.5	360	2	Q9KLT9	Q9KLT9 vibrio chol
14	97.5	10.4	231	8	Q9Z2Z5	Q9Z2Z5 littorina s
15	97.5	10.4	360	2	Q9KM39	Q9KM39 vibrio chol
16	97.5	10.4	565	8	Q9MR91	Q9MR91 loligo blue
17	97.5	10.4	566	8	Q9MJ59	Q9MJ59 loligo blue
18	96.5	10.2	485	2	Q9PG12	Q9PG12 xyella fas
19	95.5	10.1	371	14	Q9YLF8	Q9YLF8 human herpe

20	95	10.1	433	2	Q9S4G6	Q9S4G6 leptospira
21	95	10.1	921	10	Q9SGH9	Q9SGH9 arabidopsis
22	95	10.1	1320	5	Q27930	Q27930 drosophila
23	94.5	10.0	348	8	Q9TDB7	Q9TDB7 cynopocilu
24	94.5	10.0	1228	10	Q9SGG3	Q9SGG3 arabidopsis
25	94	10.0	510	2	Q9ZDE9	Q9ZDE9 rickettsia
26	93.5	9.9	393	3	Q9LAR8	Q9LAR8 bacillus ce
27	93.5	9.9	587	5	Q9VGP3	Q9VGP3 drosophila
28	93.5	9.9	870	5	Q9XND4	Q9XND4 caenorhabdi
29	93	9.9	487	2	Q67873	Q67873 aquifex aeo
30	92.5	9.8	303	2	Q34706	Q34706 bacillus su
31	92.5	9.8	395	2	Q44575	Q44575 acetobacter
32	92	9.8	407	2	Q65504	Q65504 aquifex aeo
33	91.5	9.7	267	2	Q9RNL1	Q9RNL1 zymomonas m
34	91.5	9.7	419	10	Q9MDY1	Q9MDY1 arabidopsis
35	91.5	9.7	452	2	Q9ZDB9	Q9ZDB9 rickettsia
36	91	9.7	273	2	Q66090	Q66090 lactococcus
37	91	9.7	313	8	Q99816	Q99816 ixodes hexa
38	91	9.7	425	8	Q9MD45	Q9MD45 schistosoma
39	91	9.7	435	2	Q9ZK40	Q9ZK40 helicobacte
40	91	9.7	563	2	Q56876	Q56876 yersinia en
41	91	9.7	643	1	Q29273	Q29273 archaeoglob
42	90.5	9.6	442	2	Q25239	Q25239 helicobacte
43	90	9.6	274	8	Q36360	Q36360 drosophila
44	90	9.6	381	14	Q89558	Q89558 human herpe
45	90	9.6	517	8	Q9TAK5	Q9TAK5 caeteria r

ALIGNMENTS

RESULT 1
ID Q9V235 PRELIMINARY; PRT; 175 AA.
AC Q9V235;
DT 01-MAY-2000 (TREMREL.13, Created)
DT 01-MAY-2000 (TREMREL.13, Last sequence update)
DT 01-JUN-2000 (TREMREL.14, Last annotation update)
DE HYPOTHEICAL 19.1 KDA PROTEIN.
GN PAB0162.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Pyrococcus.
OX NCBI_Taxid=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAT;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ248283; CAB49163.1; -
KW Hypothetical protein.
SQ SEQUENCE 175 AA; 19133 MW; 19ABA60C3DC15F79 CRC64;

Query Match 17.4%; Score 163.5; DB 1; Length 175;
Best Local Similarity 34.8%; Pred. No. 0.00013;
Matches 55; Conservative 28; Mismatches 64; Indels 11; Gaps 5;
QY 24 MIALA-----ISMSSWENINWNAISDGHAVKSSVAFIFNIGLAIGILIVIGRLNLY 77
DB 10 ITALAGVFAYFLNRNDWMSITENASIDGR-VGLPNMNVNMGILIGLLLGAMRLX 68
DB 69 KSKNPGMAIYVGVSVFLVIGVPEGETEW-HYEVSMGFEVSMFAMLLISISILKSRIG 127
QY 78 SMSRYKSLIISMGVFLNLICVPEDEYGMHFLVSUFLSILA--YFIAISILDKSMA 135
DB 136 VL-LIIGHIAMVYVLPASEIPRGAIPELAVESFLP 172
DB 128 FLGLAVFCVSLPLALFSLKAFSGVAVAETVSILAEPLF 165
RESULT 2

RA Heldelberg J.F., Eelsen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettein H., Richardson D.,
 RA McDonald L., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
 RT *cholerae*.";
 RL Nature 406:477-483(2000).
 DR EMBL: AF004215; AAF94519.1; -
 DR TIGR: VCJ361; -
 SQ SEQUENCE 401 AA; 44207 MW; 76923D203A8C113A CRC64;

Query Match 11.2%; Score 105.5; DB 2; Length 401;
 Best Local Similarity 20.1%; Pred. No. 2.8;
 Matches 46; Conservative 38; Mismatches 84; Indels 61; Gaps 7;

QY 6 HSWMDMIFILSF-SFPLTMTALISMSSWFNINNALSDI-----GHA 48
 DB 15 HSKHTNLNLPFRSVVQGLAVCLAFELTYNNALTNESGIANCFLEQTAFG 74
 QY 49 VKSSVAP-----IFNGLAIGLITVGLRLNYSRVSGLSI 88
 DB 75 ISQSLIPDEFTSYGRFVLGLNTALVSVGLIATLIGLIGLARLSMWLISRLAAV 134
 QY 89 SMGFVLNIGVEDEYGMHFLVSVL-----FELSIAYFAISILDKS--WI 134
 DB 135 YIEFFRNL-PLLTQFFRYFVVLQALPSARSMLGEMFYINIGLYAKTFESGSIW 193
 QY 135 AVLLIIGHIAMVYLHF-----ASEIPRGAIPPELLAVSEFLPEYI 174
 DB 194 LVALLIAGIACVLSIMATNRQLGQOTPMGRITLLCVFEPVLYVL 242

RESULT 9
 Q9RX06 PRELIMINARY; PRT; 373 AA.
 AC Q9RX06;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DE 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE CONSERVED HYPOTHETICAL PROTEIN.
 GN DR0252.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
 OX NCBI_Taxid=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R1;
 RA MEDLINE=20036896; Pubmed=10567266;
 RA White O., Eelsen J.A., Heldelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwin M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Uterback T., Zaleski C.,
 RA Mekalanos J.K., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome Sequence of the Radioresistant Bacterium *Deinococcus*
 RT *radiodurans* R1.";
 RL Science 286:1571-1577(1999).
 DR EMBL: AE001886; AAF09838.1; -
 DR TIGR: DR0252; -
 DR INTERPRO: IPR002549; -
 DR PRAM: PR01594; UPF0118; 1
 SQ SEQUENCE 373 AA; 40283 MW; 50203FE29051E67 CRC64;

Query Match 11.1%; Score 104.5; DB 2; Length 373;
 Best Local Similarity 27.4%; Pred. No. 3.1;
 Matches 43; Conservative 29; Mismatches 44; Indels 41; Gaps 9;

QY 30 SMSWF-NINNALSDIGHAVKSSVAFINLGAIGI-----LIVIGLRNLSMSRVK 83
 DB 142 ALRPMIANIGNNVIPIIOLNAL-SSTGLFNLNLSIGVIGVLLLSVIMADYRVN 200
 QY 84 GSLIISMG-----VELNIGVEDEYGMHFLVSVLFLSIIVAFISILDKSMIA 135
 DB 201 ASLRAFRPMPQPVLEFSDSLIGT-AVGGYVRGQL-----LIAAFIGIFV---WLG 247
 QY 136 VLLIIGHIAMVYLHFASIPRGAIPPELLAVSEFLPF 172
 DB 248 -LTIIVG-----IFSAALIGFLAGAFNIPVY 271

RESULT 10
 Q9KNN1 PRELIMINARY; PRT; 600 AA.
 AC Q9KNN1;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE THIOI:DISULFIDE INTERCHANGE PROTEIN DSBD.
 GN VC2701.
 OS *Vibrio cholerae*.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; *Vibrio*.
 OX NCBI_Taxid=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;
 RA MEDLINE=20406833; Pubmed=10952301;
 RA Heldelberg J.F., Eelsen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettein H., Richardson D.,
 RA McDonald L., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
 RT *cholerae*.";
 RL Nature 406:477-483(2000).
 DR EMBL: AE004336; AAF95841.1; -
 DR TIGR: VC2701; -
 SQ SEQUENCE 600 AA; 65209 MW; 23D01D76F730249C CRC64;

Query Match 11.0%; Score 103.5; DB 2; Length 600;
 Best Local Similarity 22.5%; Pred. No. 5.5;
 Matches 42; Conservative 40; Mismatches 82; Indels 23; Gaps 7;

QY 5 KHSWMDMIFILSF-SFPLTMTAL-AISMSSWFNINNALSDIGHAVKSSVAFINLGLAI 63
 DB 271 QHPYVLMLGLSVLFAVALSMFGVSLDLPSCVQFTWLSNAQCG--GSLPGVFAAGAIS 328
 QY 64 G-----GILYIVGLRLNLSRVSGLSIISMGVFLNLIVF-----DEYVGW 106
 DB 329 GLVCSPTTAPLGSALLVVAQGLTGAVLVALAIGMGIPLLVAVFGKILLPKAGNW 388
 QY 107 IHFVSVLFLFSIA-VEIAISILDKSMIAVL-LIIGHIAM-VYLHFASIPRGAIPEL 163
 DB 389 MERKKTLEGFVLLAAPFLFLERIVPERFSSVLSALGLAFAFGMLHYHKNLSLPRGKQSL 448
 QY 164 LAVESFL 170
 DB 449 IGIIVAIL 455

RESULT 11
 Q9VGYO PRELIMINARY; PRT; 293 AA.
 AC Q9VGYO;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE CG4073 PROTEIN.

GN CG4073.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Phryganea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 OX NCBI_TaxID=7227;
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang O., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-O., Andrews-Planck C., Baldwin D.,
 RA Bailly R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegym C.,
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Jansel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleby J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupel M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003688; AAF54542.1;
 DR FLYBASE: FBgn0037827; CG4073.
 SQ SEQUENCE 293 AA; 32710 MW; 7B099E726C6916D0 CRC64;

Query Match 10.9%; Score 102.5; DB 5; Length 293;
 Best Local Similarity 24.5%; Pred. No. 3.5;
 Matches 49; Conservative 32; Mismatches 62; Indels 57; Gaps 8;

QY 5 KH-SWDMITFIISFSPPLTMIATLAISSSWFNINNALSDLGAAVSSVAPINLGLAI 63
 DB 73 KHLGMYVILHVAIVLASASAVIOLVLR-----CNKFGELISPSVPSFWLLAV 121
 QY 64 GGLI-----VIYGLRLYSMSRYKG-----SLIISMGVFL 94
 DB 122 GCVLIMAVYILANOCPCNGLLATVIEVYIFVNCRRARSLMAMGVLSLVALLANVL 181
 QY 95 NLIGVDEYVGMIFLVSVLEFSLIAFIAISITLDRKSWIAVLLIGHIAMYLFHASEI 154
 DB 182 YLMG-----VYLPLKILPGSIFM--IVLFPCIALV-----VINGNRYI 225
 QY 155 PRGAIPPELLAVSFLLPYI 174
 DB 226 MRYVSNVSLIYVASLIIFETI 245

RESULT 12
 ID 09XMT1 PRELIMINARY; PRT; 365 AA.
 AC 09XMT1;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE ORF365.
 OS Tetrahymena pyriformis.
 OG Mitochondrion.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
 OC Tetrahymena; Tetrahymena.
 OX NCBI_TaxID=5908;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Edqvist J., Burger G., Gray M.W.;
 RT "Expression of mitochondrial protein-coding genes in Tetrahymena pyriformis."
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Burger G., Zhu Y., Littlejohn T.G., Greenwood S.J., Schnare M.N.,
 RA Lang B.F., Gray M.W.;
 RT "Complete sequence, gene content and organization of the mitochondrial genome of Tetrahymena pyriformis. Comparison with Paramecium aurelia RT mitochondrial DNA."
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF160864; AAD41932.1;
 DR INTERPRO: IPR001064;
 DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
 KW Mitochondrion.
 SQ SEQUENCE 365 AA; 43578 MW; 04F18A0D1252D74 CRC64;

Query Match 10.8%; Score 101.5; DB 8; Length 365;
 Best Local Similarity 24.3%; Pred. No. 4.9;
 Matches 37; Conservative 27; Mismatches 43; Indels 45; Gaps 7;

QY 11 MITTILSFPLT-----MIALAISMSWFNINNALSDL-----GHAVKSSVAPI 56
 DB 211 MIVSIFVIFLITGTPTMSVNFISNNQINLNKIALIFIAIILIGHIIRKIGIAP 270
 QY 57 FNLGLAIGLIIIVGLRNLYSWSRVKSGSLISGVLNLIGVDEYVGMIFLVSVLEF 116
 DB 271 -----QTKIIRYK-----LPFLSI--FYTFYFLVFLFF 301
 QY 117 LSIAYFLAISIDKSWIAVLLIGHIAMYL 148
 DB 302 SLIFVYILS-ALINYYWL-LLLVSIIGIYI 331
 RESULT 13
 ID 09KIL29 PRELIMINARY; PRT; 360 AA.
 AC 09KIL29;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE PEPTIDE ABC TRANSPORTER, PERMEASE PROTEIN, PUTATIVE.
 GN VCA0590.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Niernan W.C., White O.,

RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
 RT cholerae";
 RL Nature 406:477-483(2000).
 DR EMBL: AE004389; AAF96492.1; -
 DR TIGR: VCA0550; -
 SQ SEQUENCE 360 AA; 40400 MW; FC19A3822D26E2A CRC64;

Query Match 10.5%; Score 98.5; DB 2; Length 360;
 Best Local Similarity 25.8%; Pred. No. 7.8;
 Matches 34; Conservative 29; Mismatches 52; Indels 17; Gaps 6;

QY 1 MSLNKSHWMDIIFLFSFPLTMTALAISSSWFNNALSDLGHAHVSAPIFNLG 60
 DB 131 VSIISGLMSTLITLII--SIPLG--ISKAIHIGSRDVSNAVIGVAP-----G 178
 QY 61 LAIGGILIVIGLRNLYSWSRVKSGSLITSMGVFLNLIGFDEYVGMHFLVSLFELSII 120
 DB 179 FLFAIILILFASGVNFSMPFLRGIVSDNFASLPMYQGVLD--YFW--HLTLPTL--AMVI 233
 QY 121 AYFAISILDKS 132
 DB 234 GGFATLSMLTRN 245

RESULT 14
 ID 092Z05 PRELIMINARY; PRT; 231 AA.
 AC 092Z05;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE AMP SYNTHASE A CHAIN (EC 3.6.1.34).
 GN ATPASE 6.
 OS *Litorina saxatilis*.
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;
 OC Neotaenioglossa; Littorinidae; Littorinidae; Littorina.
 OX NCBI_TaxID=31220;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wilding C.S., Mill P.J., Grahame J.;
 RT "Partial sequence of the mitochondrial genome of *Littorina saxatilis*:
 RT relevance to gastropod phylogenetics";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL. IT MAY PLAY A
 CC DIRECT ROLE IN THE TRANSDUCTION OF PROTONS ACROSS THE MEMBRANE
 CC (BY SIMILARITY).
 CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
 CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 CC HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
 DR EMBL: AJ132137; CAA10596.1; -
 DR INTERPRO: IPR000568; -
 DR PFAM: PF00119; ATP-synt_A; 1.
 DR PRINTS: PR00123; ATPASEA.
 DR PROSITE: PS00449; ATPASE_A; 1.
 DR Mitochondrion; Hydrogen ion transport; CF(0); Transmembrane.
 KW SEQUENCE 231 AA; 25478 MW; 7EEAF487BC53A67A CRC64;

Query Match 10.4%; Score 97.5; DB 8; Length 231;
 Best Local Similarity 24.2%; Pred. No. 6.2;
 Matches 53; Conservative 34; Mismatches 63; Indels 69; Gaps 12;

QY 2 SLNKSWMMDIIFLFSFPLTMTALAISSSWFN--INNALSGLHGVASVAPIFNL 59
 DB 8 SFDDNNQVPSLIVLMAFLVIT--LIFSSSYWTGPRMAPIISLFKDTISSQVFRSF-- 64
 QY 60 GLAIGILIVIGLRNLYSWSRVKSGSLITSMGVFLNLIGFDEYVGMHFLVSLFELSII 105

DB 65 GLSLGGLFNLVITGL-----FLFLFLNLAGLIPYFSPSHLAVSLGLP 110
 QY 106 -WIFLVSVLEF--LSITAFI-----AISLDKS-WIANVLL-----IG 141
 DB 111 LMLSLISLIGTFENPSSVYAGLLPMGAPAPLNPFLVLESIVLPTTLISVRLMANSAG 170
 QY 142 HIAMV---YLHFASEIPRGA-----AIPELLAVFSF 169
 DB 171 HIVLITGNVMTSLFMP-GAFSMLLISITQVLYTTFEF 208

RESULT 15
 ID 09KM39 PRELIMINARY; PRT; 360 AA.
 AC 09KM39;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE HYPOTHETICAL PROTEIN VCA0550.
 GN VCA0550.
 OS *Vibrio cholerae*.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; *Vibrio*.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dracoli I., Sellers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
 RT cholerae";
 RL Nature 406:477-483(2000).
 DR EMBL: AE004386; AAF96452.1; -
 DR TIGR: VCA0550; -
 KW Hypothetical protein.
 SQ SEQUENCE 360 AA; 40265 MW; 999D1FEB2FD08867 CRC64;

Query Match 10.4%; Score 97.5; DB 2; Length 360;
 Best Local Similarity 25.9%; Pred. No. 9.2;
 Matches 45; Conservative 29; Mismatches 49; Indels 51; Gaps 10;

QY 8 WMDIIFLFSFPLTMTALAI--SSSW-----FNINNALSDLGHAHVSAPIFNL 49
 DB 197 WQDVVISIMGCFEFLSLILLYGVYWTDSFMAALAVFNAFLMLFNLPLVLPDGGHIL 256
 QY 50 KSSVAPIFNLGAILIVIGLRNLYSWSRVKSGSLITSMGVFLNLIGFDEYVGMHFLVSLFELSII 105
 DB 257 KS---VSPSMKMGVILCVAILGII-----ALVSTGLFLFGLLGMGLDVFVE 305
 QY 106 W-----IHFL-----VSVLFSLIAVFAI-----SLDKSWIAVLLIIG 141
 DB 306 WRQRHSHLPLNRYAQNVSITWIFALVSGILAITIGFASIGDILLSLPLIIG 359

Search completed: May 30, 2001, 15:25:50
 Job time: 120 sec

THIS PAGE BLANK (USPTO)